ESTIMATING THE NUMBER OF BLOCK BOUNDARIES FROM DIAGONAL BLOCKWISE MATRICES WITHOUT PENALIZATION

V. BRAULT, M. DELATTRE, E. LEBARBIER, T. MARY-HUARD, C. LÉVY-LEDUC

ABSTRACT. In computational biology, numerous recent studies have been dedicated to the analysis of the chromatin structure within the cell by two-dimensional segmentation methods. Motivated by this application, we consider the problem of retrieving the diagonal blocks in a matrix of observations. The theoretical properties of the least-squares estimators of both the boundaries and the number of blocks proposed by Lévy-Leduc et al. [2014] are investigated. More precisely, the contribution of the paper is to establish the consistency of these estimators. A surprising consequence of our results is that, contrary to the one-dimensional case, a penalty is not needed for retrieving the true number of diagonal blocks. Finally, the results are illustrated on synthetic data.

1. INTRODUCTION

Detecting change-points in one-dimensional signals is a very important task which arises in many applications, ranging from EEG (Electroencephalography) to speech processing and network intrusion detection, see Basseville and Nikiforov [1993], Brodsky and Darkhovsky [2000], Tartakovsky et al. [2014]. The aim of such approaches is to split a signal into several homogeneous segments according to some quantity. A large literature has been dedicated to the change-point detection issue for one-dimensional data. This problem may also have several applications when dealing with two-dimensional data. One of the main situations in which this problem occurs is the detection of chromosomal regions having close spatial location in the nucleus of a cell. Detecting such regions provides valuable insight to understand the influence of chromosomal conformation on cell functioning. More precisely, we will consider the problem of identifying the so-called *cis*-interactions between regions of a chromosome. In this context, n locations spatially ordered along a given chromosome are considered, the goal being to find clusters of adjacent locations that strongly interact. The elements $Y_{i,j}$ of a data matrix $\mathcal{Y}$ will then correspond to the interaction level between locations i and j of a chromosome, which can be measured using the recently developed HiC technologies, see Dixon et al. [2012]. In this application, the signal - and consequently the data matrix - exhibits a

arXiv:1506.03198v2 [math.ST] 11 Mar 2016

strong structure: one should observe high signal levels within blocks of locations along the matrix diagonal, and a signal that is close to some (low) baseline level everywhere else.

As shown in Lévy-Leduc et al. [2014], the identification of *cis*-interactions can be cast as a segmentation problem, where the goal is to identify diagonal blocks (or regions) with homogeneous interaction levels. Thanks to the spatial repartition of these regions along the diagonal, the two-dimensional segmentation of the data matrix actually boils down to a particular one-dimensional segmentation. The dynamic programming algorithm originally proposed by Bellman [1961] is well-known to provide the exact solution of the one-dimensional segmentation issue in the least-squares sense. Therefore we benefit from the data structure by avoiding both the computational burden and the approximation errors that come with heuristic methods used to solve the complex generic problem of two-dimensional segmentation.

While being able to handle large interaction data matrices from an algorithmic point of view, model selection (*i.e.* selecting the number of blocks K) remains an open question when dealing with such data. This is contrasted with the problem of one-dimensional signal segmentation, for which the properties of the estimators have been largely addressed for instance in Boysen et al. [2009], Lavielle and Moulines [2000], Yao and Au [1989]. In these approaches, the number of change-points is usually performed thanks to a Schwarz-like penalty $\lambda_n K$ where λ_n is often calibrated on data, as in Lavielle [2005] and Lavielle and Moulines [2000], or a penalty $K(a + b\log(n/K))$ as in Lebarbier [2005] and Massart [2004], where a and b are data-driven as well.

The goal of the present paper is to prove the consistency of the estimators of both the boundaries and the number of blocks obtained by minimizing the (slightly modified) least-squares criterion proposed by Lévy-Leduc et al. [2014]. The proof relies on the strong structure of the data which is of great help for the model selection issue and for the algorithmic aspects. More precisely, we will prove that the *non-penalized* least-squares estimators of the number of blocks is consistent.

The paper is organized as follows: Section 2 introduces the modeling of the data and the definition of the least-squares estimators that will be considered throughout the article. The theoretical properties of the estimators are derived in Section 3 and illustrated on synthetic data in Section 4. A discussion is given in Section 5. The technical aspects of the proofs are detailed in Section 6 and in the supplementary material.

2. Statistical Framework

2.1. **Modeling.** Let us consider $\mathcal{Y} = (Y_{i,j})_{1 \leq i,j \leq n}$, a symmetric matrix of random variables. Because of the symmetry, we shall focus on its upper-triangular part denoted by

$\mathbf{Y} = (Y_{i,j})_{1 \leq i \leq j \leq n}$ where the $Y_{i,j}$ will be assumed to be independent and such that

$$Y_{i,j} = \mathbb{E}[Y_{i,j}] + \varepsilon_{i,j} = \mu_{i,j} + \varepsilon_{i,j}, \ 1 \leq i \leq j \leq n. \tag{1}$$

The $\varepsilon_{i,j}$ satisfy the following assumption:

(**A1**) The $\varepsilon_{i,j}$ are assumed to be centered, i.i.d. and such that there exists a positive constant β such that for all $\nu \in \mathbb{R}$,

$$\mathbb{E}[e^{\nu \varepsilon_{11}}] \leq e^{\beta \nu^2}.$$

We shall moreover assume that the matrix of means $(\mu_{i,j})_{1 \leq i \leq j \leq n}$ is block diagonal. More precisely, let $\boldsymbol{\tau}^\star = (\tau_0^\star, \tau_1^\star, \ldots, \tau_{K^\star}^\star)$ be a vector of break fractions such that $0 = \tau_0^\star < \tau_1^\star < \cdots < \tau_{K^\star}^\star = 1$. In what follows, the break fractions are fixed quantities: neither their number nor their positions change when n grows. The parameters $\mu_{i,j}$ are such that

$$\begin{aligned} \mu_{i,j} &= \mu_k^\star \quad \text{if } (i,j) \in D_k^\star, \ k = 1, \ldots, K^\star, \\ &= \mu_0^\star \quad \text{if } (i,j) \in E_0^\star, \end{aligned} \tag{2}$$

where the (half) diagonal blocks $D_k^\star$ $(k = 1, \ldots, K^\star)$ are defined as follows,

$$D_k^\star = \{(i,j) : t_{k-1}^\star \leq i \leq j \leq t_k^\star - 1\}, \tag{3}$$

where $t_k^\star = [n\tau_k^\star] + 1$ are thus such that $1 = t_0^\star < t_1^\star < \cdots < t_{K^\star}^\star = n + 1$, $[x]$ denoting the integer part of x. They stand for the true block boundaries and $K^\star$ corresponds to the true number of blocks. In Equation (2), $E_0^\star$ corresponds to the set of positions lying outside the diagonal blocks:

$$E_0^\star = \{(i,j) : 1 \leq i \leq j \leq n\} \cap (\cup D_k^\star)^C, \tag{4}$$

where A^C denotes the complement of set A. An example of such a matrix is displayed in Figure 1 (left). The following will also be assumed for the true block sizes:

(**A2**) For all ℓ, one has

$$0 < \Delta_{\boldsymbol{\tau}}^\star = \min_{k \in \{1, \ldots, K^\star\}} \left| \tau_k^\star - \tau_{k-1}^\star \right| \leq |\tau_{\ell+1}^\star - \tau_\ell^\star| \leq c,$$

where $c \in (0, 1)$ is a known constant.

Moreover the $\mu_k^\star$ satisfy the following assumption:

(**A3**) $$\underline{\lambda}^{(0)} = \min_{1 \leq k \leq K^\star} |\mu_k^\star - \mu_0^\star| > 0.$$

2.2. **Inference.** In this framework, the inference consists in estimating both the number of blocks and the true break fraction vector $\boldsymbol{\tau}^\star$ (or equivalently the true boundary vector $\mathbf{t}^\star$). One strategy would be to use the following least-squares criterion:

$$\widehat{\mathbf{t}}_K^{\mathrm{LS}} \in \underset{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}}{\mathrm{Argmin}} \left\{ \left[\sum_{k=1}^K \sum_{(i,j) \in D_k} \left(Y_{i,j} - \overline{Y}_{D_k} \right)^2 \right] + \sum_{(i,j) \in E_0} \left(Y_{i,j} - \overline{Y}_{E_0} \right)^2 \right\}, \tag{5}$$

where $\overline{Y}_{\mathcal{D}}$ is the empirical mean of the $Y_{i,j}$ when the indices (i,j) belong to $\mathcal{D}$, D_k and E_0 are defined as in (3) and (4) except that $\mathbf{t}^\star$ is replaced by $\mathbf{t}$, and K is the considered number of segments – $K^\star$ being unknown in practice. Moreover,

$$\mathcal{A}_{n,K}^{\Delta_n} = \{ \mathbf{t} = (t_0, \ldots, t_K) : t_0 = 1 < t_1 < \ldots < t_K = n+1$$

$$\text{and } \forall 1 \leq k \leq K, \ n\Delta_n \leq t_k - t_{k-1} < cn \} \tag{6}$$

is the set of admissible segmentations, where Δ_n denotes a positive sequence.

However, thanks to (A2), one can derive an unbiased estimator of $\mu_0^\star$ using the upper-right triangle part of the matrix $\mathcal{Y}$ denoted G_{01} and defined by

$$G_{01} = \{(i,j) : 1 \leq i \leq n^0, (n - n^0 + 1) \leq j \leq n\} \quad \text{with } n^0 = [(1-c)n]. \tag{7}$$

Indeed the intersection between the blocks D_k and G_{01} will always be empty. Thus, we can split $E_0^\star$ into two disjoint sets $G_{00}^\star$ and G_{01} (see the right part of Figure 1) as follows,

$$E_0^\star = G_{00}^\star \cup G_{01}. \tag{8}$$

Consequently, we will consider the following slightly modified least-squares criterion:

$$\widehat{\mathbf{t}}_K \in \underset{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}}{\mathrm{Argmin}} \ Q_n^K(\mathbf{t}), \tag{9}$$

where

$$Q_n^K(\mathbf{t}) = \left\{ \left[\sum_{k=1}^K \sum_{(i,j) \in D_k} \left(Y_{i,j} - \overline{Y}_{D_k} \right)^2 \right] + \sum_{(i,j) \in E_0} \left(Y_{i,j} - \overline{Y}_{G_{01}} \right)^2 \right\}. \tag{10}$$

Lastly, we will consider the following estimator of $K^\star$:

$$\widehat{K} = \underset{1 \leq K \leq K_{\max}}{\mathrm{Argmin}} \ Q_n^K \left(\widehat{\mathbf{t}}_K \right), \tag{11}$$

where $\widehat{\mathbf{t}}_K$ is defined in (9) and $K_{\max}$ is the maximal number of blocks considered.

Criterion (11) based on (10) has been proposed by Lévy-Leduc et al. [2014]. The goal of our paper is to validate this latter approach theoretically. Note that the main difference between (5) and (10) is the estimation of $\mu_0^\star$ that is independent from the segmentation, since G_{01} is fixed. Hence, $\mu_0^\star$ can be estimated prior to the optimization of the criterion (10).
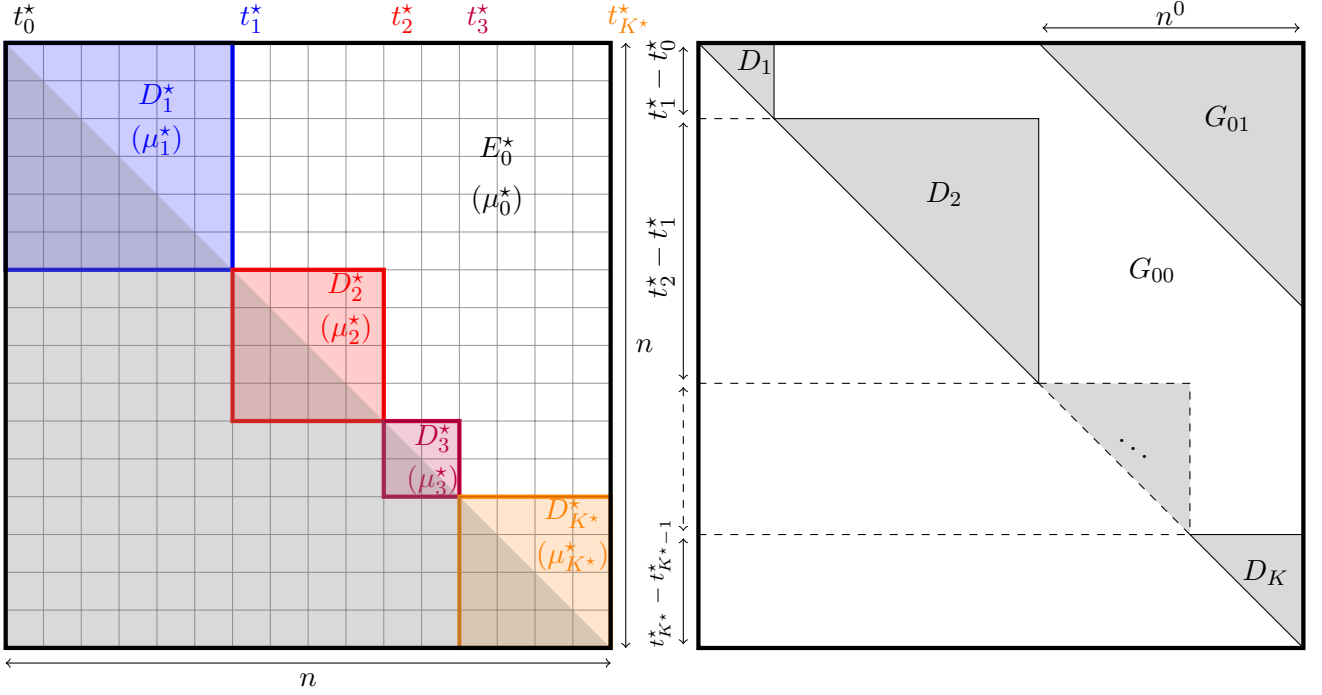
FIGURE 1. Left: Example of a matrix $(\mu_{i,j})$ with $n = 16$ and $K^\star = 4$. Right: Illustration of the notations used in the estimation criterion.

As a consequence, this optimization can be performed by using the dynamic programming algorithm as explained in Lévy-Leduc et al. [2014].

3. THEORETICAL RESULTS

The goal of this section is to derive the consistency of $\widehat{K}$ and $\hat{\boldsymbol{\tau}}$. To prove these results, we shall need the following assumption on Δ_n:

(**A4**) $$\Delta_n \frac{\sqrt{n}}{(\log n)^{1/4}} \underset{n \to +\infty}{\longrightarrow} +\infty \quad \text{and} \quad \Delta_n \leq \Delta_{\boldsymbol{\tau}}^\star, \text{ for large enough } n.$$

Theorem 1. *Let $Y_{i,j}$ be defined by (1). Assume that (A1), (A2), (A3) and (A4) hold. Then $\widehat{K}$ defined in (11) is such that:*

$$\mathbb{P}\left(\widehat{K} \neq K^\star\right) \longrightarrow 0, \text{ as } n \to +\infty. \tag{12}$$

Remark 1. Observe that, contrary to classical statistical frameworks, $\widehat{K}$ is a consistent estimator of $K^\star$ even if it is obtained without any penalization.

Remark 2. In Theorem 1, the estimator $\widehat{K}$ is defined as the minimizer of $Q_n^K(\widehat{\mathbf{t}}_K)$ where $\widehat{\mathbf{t}}_K$ is obtained by minimizing $Q_n^K(\mathbf{t})$ over the set $\mathcal{A}_{n,K}^{\Delta_n}$. If we are only interested in proving that

$\mathbb{P}(\widehat{K} < K^\star) \to 0$, the minimization can be performed on the set $\mathcal{A}_{n,K}^{1/n}$ instead of $\mathcal{A}_{n,K}^{\Delta_n}$, $i.e$ without any constraint on the minimal distance between two consecutive change-points, see Lemma 1 (i) below and Lemmas 2, 3 and 4, which are given in Section 6.

Remark 3. Theorem 1 is valid under (A2) which implies that the number of observations within each segment increases linearly with n, since $t_k^\star = [n\boldsymbol{\tau}_k^\star] + 1$. This assumption could be alleviated by assuming that $\Delta_{\boldsymbol{\tau}}^\star$ is no longer a constant. In that case, we shall need to assume that $\Delta_{\boldsymbol{\tau}}^\star n^{1/4}/(\log n)^{1/8}$ tends to infinity, as n tends to infinity.

Remark 4. The assumption $\Delta_n \gg (\log n)^{1/4}/\sqrt{n}$ of (A4) can be understood in the light of Lemma 1 (ii) and Equation (17) at the end of the proof of Theorem 1. It is required to ensure the convergence to zero of the exponential inequalities of the random parts given in Lemmas 2, 3 and 4. This assumption is only required for proving that $\mathbb{P}\left(\widehat{K} > K^\star\right)$ tends to zero as n tends to infinity. As a consequence, when the number of blocks is known ($\widehat{K} = K^\star$), the break fractions consistency is obtained in our paper when $\Delta_n = 1/n$. Such a choice is impossible in the one-dimensional segmentation framework of Lavielle and Moulines [2000] since it is required that $n\Delta_n \to +\infty$ and $\Delta_n \to 0$, as n tends to infinity, in order to obtain the break fractions consistency when the number of breaks is known.

Remark 5. In practice, c has to be chosen in order to use the top right part of the matrix of observations to estimate the parameter $\mu_0^\star$. This choice can either come from a prior biological knowledge or from a simple visualization of the data. In the case of the analysis of HiC data, the size of the interaction diagonal blocks are expected to be small compared with the size of the chromosome $i.e.$ the size of the data matrix. In this context, $c = 3/4$ can be safely chosen, as suggested in Lévy-Leduc et al. [2014]. If the value of c is misspecified, the estimator of $\mu_0^\star$ is biased. The consistency result of Theorem 1 still holds if (A3) is replaced by $\min_{1 \le k \le K^\star} \left| \mu_k^\star - \mathbb{E}(\bar{Y}_{G_{01}}) \right| > 0$.

Sketch of proof of Theorem 1. In order to prove (12), we shall prove that $\mathbb{P}\left(\widehat{K} < K^\star\right)$ and $\mathbb{P}\left(\widehat{K} > K^\star\right)$ tend to zero as n tends to infinity. Note that

$$\mathbb{P}\left(\widehat{K} < K^\star\right) \le \sum_{K=1}^{K^\star-1} \mathbb{P}\left(\widehat{K} = K\right) \text{ and } \mathbb{P}\left(\widehat{K} > K^\star\right) \le \sum_{K=K^\star+1}^{K_{\max}} \mathbb{P}\left(\widehat{K} = K\right).$$

Hence, we shall prove that for $K < K^\star$ and $K > K^\star$,

$$\mathbb{P}\left(\widehat{K} = K\right) \longrightarrow 0, \text{ as } n \to +\infty.$$

Observe that by definition of $\widehat{K}$ given in (11),

$$\mathbb{P}\left(\widehat{K} = K\right) \leq \mathbb{P}\left(\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} Q_n^K(\mathbf{t}) - \min_{\mathbf{t} \in \mathcal{A}_{n,K^\star}^{\Delta_n}} Q_n^{K^\star}(\mathbf{t}) \leq 0\right)$$

$$\leq \mathbb{P}\left(\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} Q_n^K(\mathbf{t}) - Q_n^{K^\star}(\mathbf{t}^\star) \leq 0\right),$$

since, for large enough n, $\Delta_n \leq \Delta_{\boldsymbol{\tau}}^\star$, and hence $\mathbf{t}^\star$ belongs to $\mathcal{A}_{n,K^\star}^{\Delta_n}$. Thus, we shall focus on

$$\mathbb{P}\left(\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} J_n(\mathbf{t}) \leq 0\right),$$

where

$$J_n(\mathbf{t}) = \frac{2}{n(n+1)}\left(Q_n^K(\mathbf{t}) - Q_n^{K^\star}(\mathbf{t}^\star)\right), \tag{13}$$

We shall prove in the supplementary material that

$$J_n(\mathbf{t}) = \mathrm{B}_n(\mathbf{t}) + V_n(\mathbf{t}) + W_n(\mathbf{t}) + Z_n(\mathbf{t}), \tag{14}$$

where B_n, V_n, W_n and Z_n are defined by (20), (21), (22), (23) and (24) in Section 6. In (14), B_n corresponds to the deterministic part and the other terms correspond to the random part of J_n.

The remainder of the proof is based on Lemma 1, which is proved in Section 6.2 and which provides a lower bound for the deterministic part of J_n, and on Lemmas 2, 3 and 4, given in Section 6, which provide deviation inequalities for the random terms of J_n.

Lemma 1. *Let $B_n(\mathbf{t})$ be defined by (20) and (21), then*

(i) if $K < K^\star$,

$$\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}} B_n(\mathbf{t}) \geq \frac{\lambda^{(0)^2}}{64}\left(\Delta_{\boldsymbol{\tau}}^\star\right)^4,$$

(ii) if $K > K^\star$,

$$\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} B_n(\mathbf{t}) \geq \frac{\lambda^{(0)^2}}{4}\Delta_n^2,$$

(iii) if $K = K^\star$, for all positive δ,

$$\min_{\left\{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}, \|\mathbf{t} - \mathbf{t}^\star\|_\infty > n\delta\right\}} B_n(\mathbf{t}) \geq \frac{\lambda^{(0)^2}}{32}\min\left(\Delta_{\boldsymbol{\tau}}^\star/2, \delta\right)\left(\Delta_{\boldsymbol{\tau}}^\star\right)^3, \tag{15}$$

where $\Delta_\tau^\star$ is defined in (A2), $\underline{\lambda}^{(0)}$ is defined in (A3) and $\mathcal{A}_{n,K}^{\Delta_n}$ is defined in (6). $\mathcal{A}_{n,K}^{1/n}$ is a particular case with $\Delta_n = 1/n$ and

$$\|\mathbf{t} - \mathbf{t}^\star\|_\infty = \max_{0 \leq k \leq K^\star} |t_k - t_k^\star|. \tag{16}$$

Thus,

$$\mathbb{P}\left(\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} J_n(\mathbf{t}) \leq 0\right) \leq \mathbb{P}\left(\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} [\mathrm{B}_n(\mathbf{t}) + V_n(\mathbf{t}) + W_n(\mathbf{t}) + Z_n(\mathbf{t})] \leq 0\right).$$

The right hand side (rhs) of the previous inequality is bounded by

$$\mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} V_n(\mathbf{t}) - \min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} W_n(\mathbf{t}) - \min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} Z_n(\mathbf{t}) \geq \min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} \mathrm{B}_n(\mathbf{t})\right).$$

For bounding this term we shall use Lemma 1 (ii). For $K > K^\star$, we obtain

$$\mathbb{P}\left(\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} J_n(\mathbf{t}) \leq 0\right) \leq \mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} V_n(\mathbf{t}) \geq \frac{\lambda^{(0)2}}{12}\Delta_n^2\right)$$

$$+ \quad \mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} W_n(\mathbf{t}) \geq \frac{\lambda^{(0)2}}{12}\Delta_n^2\right) + \mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}} Z_n(\mathbf{t}) \geq \frac{\lambda^{(0)2}}{12}\Delta_n^2\right). \tag{17}$$

By Lemmas 2, 3 and 4, we conclude that

$$\mathbb{P}\left(\widehat{K} = K\right) \xrightarrow[n \to +\infty]{} 0,$$

for $K > K^\star$. The case $K < K^\star$ can be proved by following the same lines.

$\square$

Remark 6. We can observe from Theorem 1 that adding a penalty term is not necessary for obtaining a consistent estimator of the number of diagonal blocks. This may be surprising since, in the one-dimensional case, it is proved in Theorem 9 of Lavielle and Moulines [2000] that a penalty term is required. More precisely, the main difference between our two-dimensional framework and the one-dimensional case is the behavior of the deterministic part of our criterion B_n: it is lower bounded whatever the value of K ($K \geq K^\star$ or $K < K^\star$), as proved in Lemma 1. On the contrary, in the one-dimensional case, a penalty term of the type $\beta_n K$ is necessary to obtain such a lower bound when $K \geq K^\star$. In the case where $K < K^\star$, a lower bound for B_n is obtained without penalization. For further details, see the proof of Theorem 9 in Lavielle and Moulines [2000].

Theorem 2. *Assume that the assumptions of Theorem 1 hold then, for all $\delta > 0$,*

$$\mathbb{P}\left(\left\|\mathbf{t}^\star - \widehat{\mathbf{t}}_{\widehat{K}}\right\|_{\mathcal{H}} > n\delta\right) \xrightarrow[n \to +\infty]{} 0, \tag{18}$$

where $\widehat{\mathbf{t}}_{\widehat{K}}$ is defined in (9) and (11) and $\|\cdot\|_{\mathcal{H}}$ denotes the Hausdorff distance defined by

$$\left\|\mathbf{t}^\star - \widehat{\mathbf{t}}_K\right\|_{\mathcal{H}} = \max\left[\max_{0 \leq k \leq K^\star} \min_{0 \leq \ell \leq K} \left|t_k^\star - \widehat{t}_\ell\right|, \max_{0 \leq \ell \leq K} \min_{0 \leq k \leq K^\star} \left|t_k^\star - \widehat{t}_\ell\right|\right].$$

Observe that (18) can be rewritten as $\mathbb{P}\left(\left\|\boldsymbol{\tau}^\star - \widehat{\boldsymbol{\tau}}_{\widehat{K}}\right\|_{\mathcal{H}} > \delta\right) \xrightarrow[n \to +\infty]{} 0$, where $\widehat{\boldsymbol{\tau}}_{\widehat{K}} = \widehat{\mathbf{t}}_{\widehat{K}}/n$.

Sketch of proof of Theorem 2. Observe that

$$\mathbb{P}\left(\left\|\mathbf{t}^\star - \widehat{\mathbf{t}}_{\widehat{K}}\right\|_{\mathcal{H}} > n\delta\right) = \mathbb{P}\left(\left\{\left\|\mathbf{t}^\star - \widehat{\mathbf{t}}_{\widehat{K}}\right\|_{\mathcal{H}} > n\delta\right\} \cap \left\{\widehat{K} \neq K^\star\right\}\right)$$

$$+ \mathbb{P}\left(\left\{\left\|\mathbf{t}^\star - \widehat{\mathbf{t}}_{\widehat{K}}\right\|_{\mathcal{H}} > n\delta\right\} \cap \left\{\widehat{K} = K^\star\right\}\right) \leq \mathbb{P}\left(\widehat{K} \neq K^\star\right) + \mathbb{P}\left(\|\widehat{\mathbf{t}}_{K^\star} - \mathbf{t}^\star\|_\infty > n\delta\right)$$

where $\|\widehat{\mathbf{t}}_{K^\star} - \mathbf{t}^\star\|_\infty$ is defined in (16) since $\|\widehat{\mathbf{t}}_{\widehat{K}} - \mathbf{t}^\star\|_\infty = \|\widehat{\mathbf{t}}_{\widehat{K}} - \mathbf{t}^\star\|_{\mathcal{H}}$ when $\widehat{K} = K^\star$. By Theorem 1, proving (18) amounts to proving that

$$\mathbb{P}\left(\max_{0 \leq k \leq K^\star} |t_k^\star - \widehat{t}_k| > n\delta\right) \to 0, \text{ as } n \to +\infty.$$

Observe that

$$\mathbb{P}\left(\max_{1 \leq k \leq K^\star} \left|t_k^\star - \widehat{t}_k\right| > n\delta\right) \leq \mathbb{P}\left(\min_{\left\{\mathbf{t} \in \mathcal{A}_{n,K^\star}^{1/n}, \|\mathbf{t} - \mathbf{t}^\star\|_\infty > n\delta\right\}} J_n(\mathbf{t}) \leq 0\right).$$

Using the same arguments as those used in the proof of Theorem 1, the proof follows from the decomposition of J_n given by (14), the lower bound (15) of Lemma 1 and the deviation inequalities for the random terms given by Lemmas 2, 3 and 4. $\qquad \square$

4. NUMERICAL EXPERIMENTS

The goal of this section is to illustrate the theoretical results obtained in Section 3. For an application of our method to real data, we refer the reader to Lévy-Leduc et al. [2014].

4.1. **Simulation framework.** We generated Gaussian diagonal block matrices according to Model (1) with $\mu_k^\star = 1$ for the $K^\star = 5$ diagonal blocks and $\mu_0^\star = 0$ for different values of n ($n \in \{500, 1500\}$). The change-point locations are $(\tau_0^\star, \ldots, \tau_5^\star) = (0, 0.07, 0.2, 0.4, 0.67, 1)$ hence $\Delta_\tau^\star = 0.07$. We shall use different values for the standard deviation σ of the $\varepsilon_{i,j}$: $\sigma \in \{1, ..., 10\}$. For each case, 500 matrices were simulated and the procedure was tested. Examples of such matrices are displayed in Figure 2 for different values of σ.

The results that are presented below have been obtained by using the R package HiCseg which is available on the CRAN. In this package, the values of Δ_n and c are fixed and equal to $2/n$ and $3/4$, respectively.
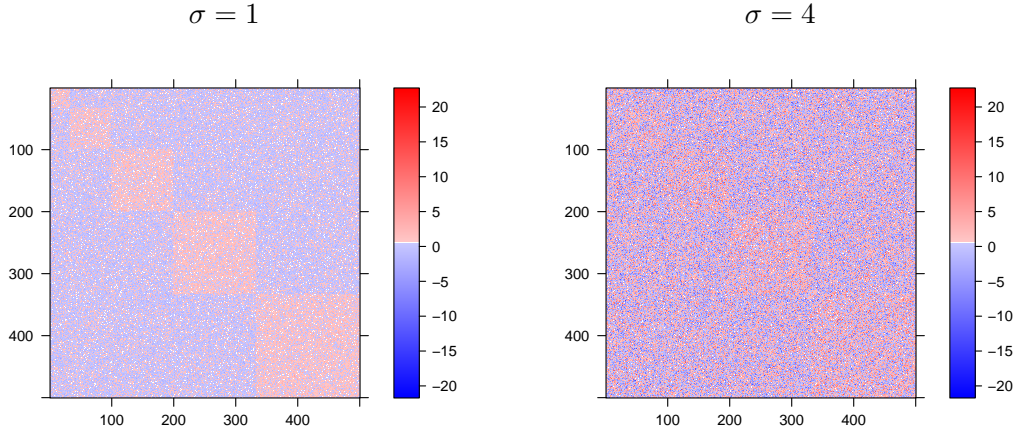
$$\sigma = 1 \qquad\qquad\qquad \sigma = 4$$

FIGURE 2. Examples of simulated matrices following Model (1) with $(\tau_0^\star, \ldots, \tau_5^\star) = (0, 0.07, 0.2, 0.4, 0.67, 1)$ and $n = 500$ for two values of σ: $\sigma = 1$ (left) and $\sigma = 4$ (right).

4.2. Statistical performance.

4.2.1. *Performance of the statistical procedure.* We first consider the problem of estimating the true number of blocks $K^\star$, and provide some insight about the consistency of our procedure without penalty, outlined in Remark 1. The median, 1st and 3rd quantiles of the estimated number of change points are displayed in Figure 3 for n in $\{500, 1500\}$ and for different values of σ.

On the one hand, we observe that for high signal to noise ratios, the true value of $K^\star$ is retrieved by our procedure. On the other hand, when the signal to noise ratio becomes very low, $K^\star$ is not properly estimated. In this situation, $K^\star$ is overestimated, which is in accordance with what occurs in the one-dimensional case where a non-penalized procedure would result in a systematic overestimation of $K^\star$. However, when n increases, the value of σ from which this overestimation occurs is unsurprisingly larger.

To illustrate the performance of our procedure in terms of the estimation of change-point location, Figure 4 displays the boxplots of the two parts of the Hausdorff distance defined by:

$$
\begin{aligned}
\left\| \mathbf{t}^\star - \widehat{\mathbf{t}}_{\widehat{K}} \right\|_{\mathcal{H}^1} &= \max_{0 \le k \le K^\star} \min_{0 \le \ell \le \widehat{K}} \left| t_k^\star - \widehat{t}_\ell \right|, \\
\left\| \mathbf{t}^\star - \widehat{\mathbf{t}}_{\widehat{K}} \right\|_{\mathcal{H}^2} &= \max_{0 \le \ell \le \widehat{K}} \min_{0 \le k \le K^\star} \left| t_k^\star - \widehat{t}_\ell \right|.
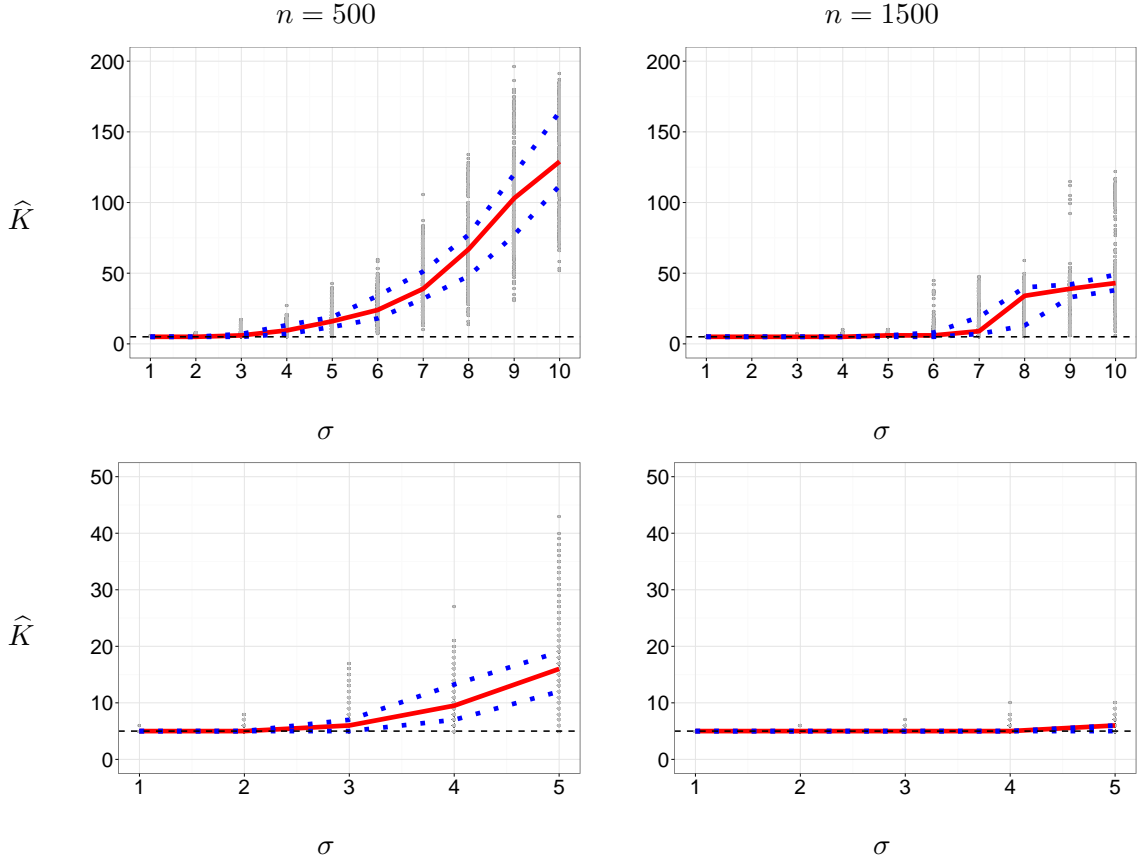\end{aligned}
\tag{19}
$$

FIGURE 3. Top: Median (plain), 1st and 3rd quartiles (dotted line) of the estimations of $K^\star = 5$ as a function of the standard deviation σ for $n = 500$ (left) and $n = 1500$ (right). The values of $\widehat{K}$ at each simulation are displayed with light grey dots. The dashed line corresponds to the true value of $K^\star$. Bottom: Same plots with the x-axis values restricted to $\{1, \dots, 5\}$.

We observe from this figure that when $K^\star$ is overestimated, the true change-points are recovered ($\|\cdot\|_{\mathcal{H}^1}$ is close to 0), the other estimated change-points being spurious ones ($\|\cdot\|_{\mathcal{H}^2}$ is large). As proved in Theorem 2, this phenomenon is less visible when n becomes large.

4.2.2. *Effect of a poor estimation of $\mu_0^\star$.* We study the behavior of our segmentation procedure when $\mu_0^\star$ is poorly estimated which may occur, for instance, when the constant c appearing in (7) is too small. To this end, we generated data in which the mean of the $n_0 \times n_0$ top right part of the observation matrix is modified, where n_0 is defined in (2). More precisely, the mean of this part is equal to $\mu_0^\star + \omega$, where $\omega \in \{0.2, 0.4, 0.6, 0.8\}$. The results are displayed in Figure 5. We can see from this figure that when the value of $\mu_0^\star + \omega$ is close to the values of
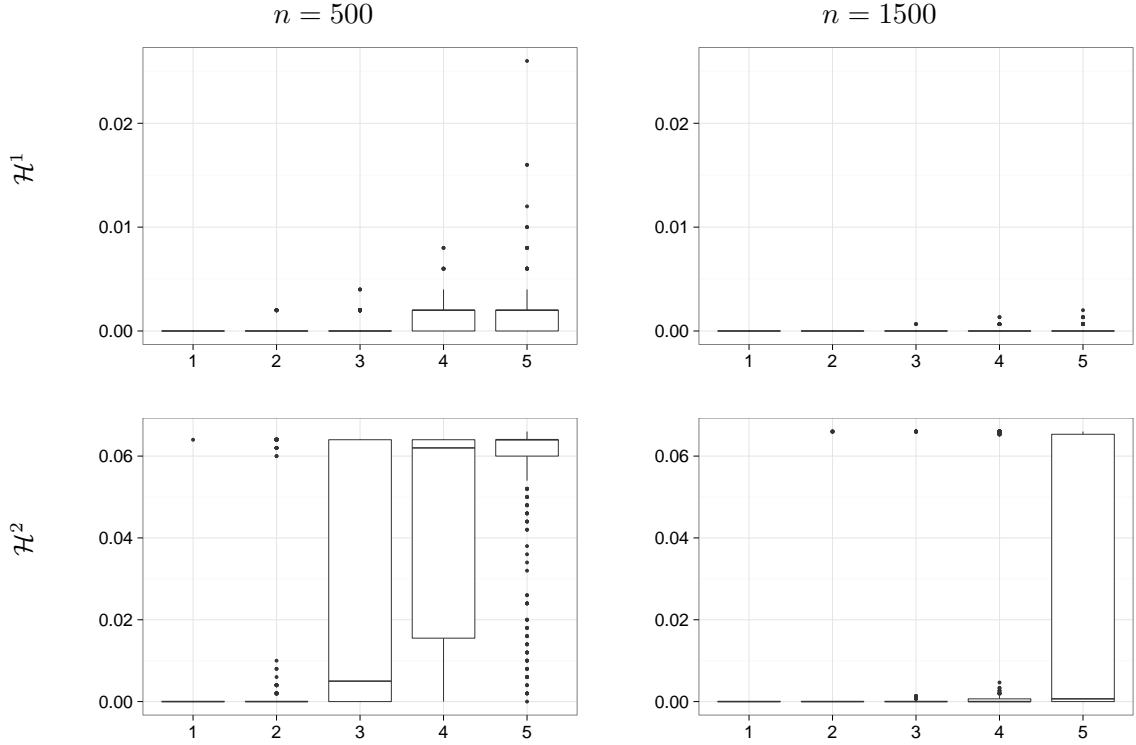
FIGURE 4.   Boxplots of the two parts of the Hausdorff distance: $\mathcal{H}^1$ (top) and $\mathcal{H}^2$ (bottom) for $n = 500$ (left) and $n = 1500$ (right). For each case, the boxplots are displayed as a function of σ.

the means of the diagonal blocks our procedure tends to overestimate $K^\star$. This phenomenon is less visible when n is large.

5. DISCUSSION

In this paper, we established that the (slightly modified) least-squares estimators for the number of blocks and their boundaries in a block diagonal matrix are consistent. Note that the obtained results are non standard in the sense that we proved that penalizing the least-squares criterion is not required to obtain a consistent estimator of the number of diagonal blocks. This has to be contrasted with the one-dimensional case, where it is well-known that a penalization is required to ensure consistency, see for instance Lavielle and Moulines [2000]. More precisely, a close look at the proof of Theorem 9 in Lavielle and Moulines [2000] shows that a penalty is required to discard models such that $K > K^\star$. This comes from the fact that in the one-dimensional setting when $K > K^\star$ the deterministic part B_n of J_n vanishes for all segmentations $\mathbf{t}$ satisfying $\|\mathbf{t}^\star - \mathbf{t}\|_{\mathcal{H}^1} = 0$ (*i.e.* for all segmentations $\mathbf{t}$ nested in the true
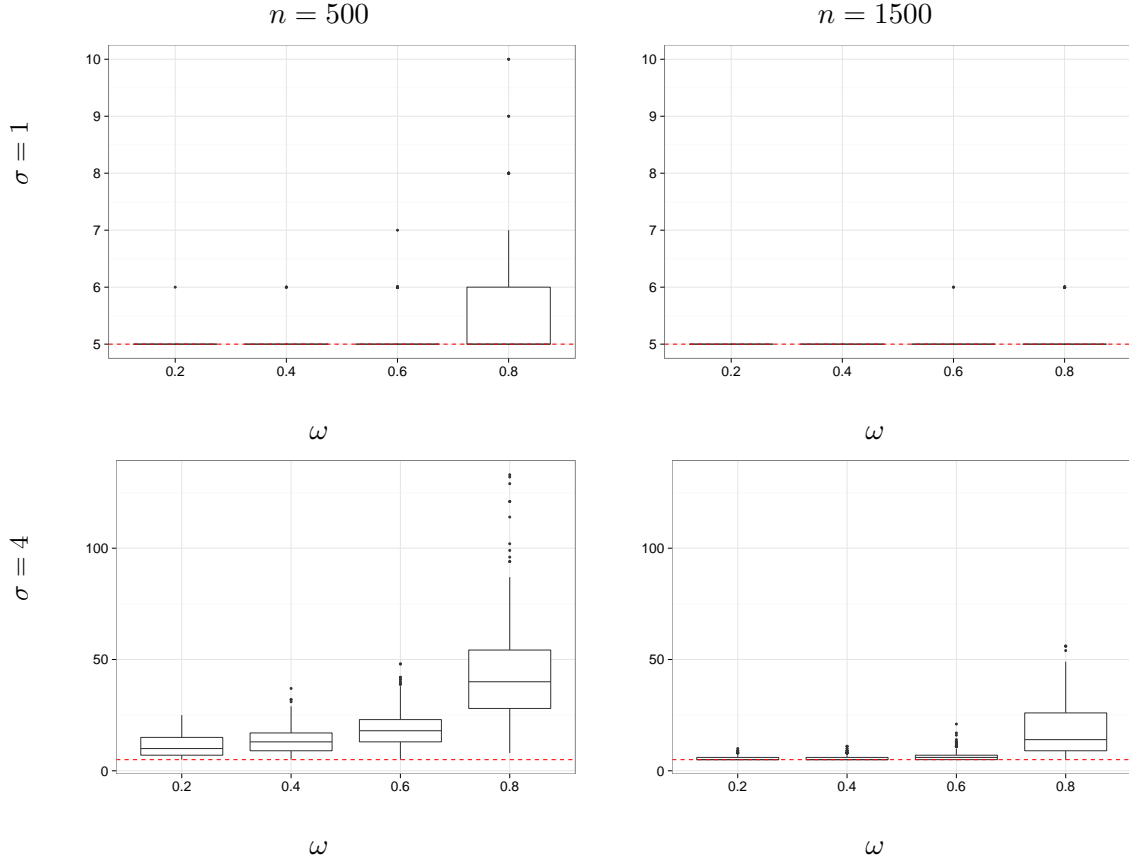
FIGURE 5.   Boxplots of $\widehat{K}$ for $\sigma = 1$ (top) and $\sigma = 4$ (bottom) for $n = 500$ (left) and $n = 1500$ (right). For each case, the boxplots are displayed as a function of ω.

segmentation $\mathbf{t}^\star$). This bias term being null, a penalty term has to be added to the criterion to compensate the stochastic deviations of the random terms in J_n. In the two-dimensional setting, the deterministic part B_n does not vanish when $K > K^\star$ –as proved in Lemma 1– ensuring consistency.

The framework that we have chosen for proving our results consists in assuming that the observations are independent and that the size of the observation matrix is large (asymptotic framework), which is adapted to the analysis of HiC experiments. From a practical point of view, the independence assumption is not always satisfied, for instance when the observation matrix is a correlation or a similarity matrix, see for example Dehman et al. [2015], Ioanna Delatola et al. [2015]. Hence, relaxing the independence assumption to retrieve diagonal block boundaries in such cases would be a natural extension of this paper. Moreover, it could be interesting to see if a penalty term needs to be added to our criterion in order to retrieve

properly the break fractions in a non asymptotic setting. This will be the subject of a future work.

6. PROOFS

6.1. Definition of B_n, V_n, W_n and Z_n.
We define hereafter B_n, V_n, W_n and Z_n which appear in (14) by:

$$B_n(\mathbf{t}) = B_n^D(\mathbf{t}) + B_n^0(\mathbf{t}), \ V_n(\mathbf{t}) = V_n^D(\mathbf{t}) + V_n^0(\mathbf{t}), \ W_n = W_n^D(\mathbf{t}) + W_n^0(\mathbf{t}), \qquad (20)$$

and

$$B_n^D(\mathbf{t}) = \frac{2}{n(n+1)} \left(\sum_{k=1}^{K} \sum_{(i,j) \in D_k} \left(\mathbb{E}\left[Y_{i,j}\right] - \mathbb{E}\left[\overline{Y}_{D_k}\right] \right)^2 \right),$$

$$B_n^0(\mathbf{t}) = \frac{2}{n(n+1)} \sum_{(i,j) \in G_{00}} \left(\mathbb{E}\left[Y_{i,j}\right] - \mathbb{E}\left[\overline{Y}_{G_{01}}\right] \right)^2, \quad (21)$$

$$V_n^D(\mathbf{t}) = \frac{2}{n(n+1)} \left[\sum_{k=1}^{K^\star} \frac{\left(\sum_{(i,j) \in D_k^\star} \varepsilon_{i,j} \right)^2}{|D_k^\star|} - \sum_{k=1}^{K} \frac{\left(\sum_{(i',j') \in D_k} \varepsilon_{i',j'} \right)^2}{|D_k|} \right],$$

$$V_n^0(\mathbf{t}) = \frac{2}{n(n+1)} \frac{1}{|G_{01}|^2} \left(\sum_{(i,j) \in G_{01}} \varepsilon_{i,j} \right)^2 \left(|G_{00}| - |G_{00}^\star| \right), \quad (22)$$

$$W_n^D(\mathbf{t}) = \frac{4}{n(n+1)} \left[\sum_{k=1}^{K^\star} \left(\sum_{(i,j) \in D_k^\star} \varepsilon_{i,j} \right) \mu_k^\star - \sum_{k=1}^{K} \left[\left(\sum_{(i',j') \in D_k} \varepsilon_{i',j'} \right) \mathbb{E}\left[\overline{Y}_{D_k}\right] \right] \right],$$

$$W_n^0(\mathbf{t}) = \frac{4}{n(n+1)} \mu_0^\star \left(\sum_{(i,j) \in G_{00}^\star} \varepsilon_{i,j} - \sum_{(i,j) \in G_{00}} \varepsilon_{i,j} \right), \quad (23)$$

$$Z_n(\mathbf{t}) = \frac{4}{n(n+1)} \frac{1}{|G_{01}|} \left(\sum_{(i,j) \in G_{01}} \varepsilon_{i,j} \right) \left[\left(\sum_{(i,j) \in G_{00}^\star} \varepsilon_{i,j} - \sum_{(i,j) \in G_{00}} \varepsilon_{i,j} \right) - \sum_{(i,j) \in G_{00}} \left(\mathbb{E}\left[Y_{i,j}\right] - \mu_0^\star \right) \right].$$
$$(24)$$

In the equations, $G_{00}^\star$ and G_{01} are defined in (8) and (7) and G_{00} has the same definition as $G_{00}^\star$ except that $\mathbf{t}^\star$ is replaced by $\mathbf{t}$.

6.2. **Proof of Lemma 1.** We shall first rewrite B_n^D and B_n^0 defined by (21). Let us first denote by

$$n_{k,\ell} = |D_k \cap D_\ell^\star|, \tag{25}$$

the number of observations that belong to the intersection of the two blocks D_k and $D_\ell^\star$ (with the convention that $D_0 = G_{00}$ and $D_0^\star = G_{00}^\star$) and

$$n_k = \sum_{\ell=0}^{K^\star} n_{k,\ell} \quad \text{and} \quad n_\ell^\star = \sum_{k=0}^{K} n_{k,\ell}.$$

Since $\mathbb{E}\left[\overline{Y}_{G_{01}}\right] = \mu_0^\star$, $G_{00} \subset \left(\bigcup_{\ell=0}^{K^\star} D_\ell^\star\right)$ and $\mathbb{E}\left[Y_{i,j}\right] = \mu_k^\star$, for all $(i,j) \in D_k^\star$, we obtain

$$
\begin{aligned}
B_n^0(\mathbf{t}) &= \frac{2}{n(n+1)} \sum_{(i,j)\in G_{00}} \left(\mathbb{E}\left[Y_{i,j}\right] - \mu_0^\star\right)^2 = \frac{2}{n(n+1)} \sum_{\ell=0}^{K^\star} \sum_{(i,j)\in G_{00}\cap D_\ell^\star} \left(\mathbb{E}\left[Y_{i,j}\right] - \mu_0^\star\right)^2 \\
&= \frac{2}{n(n+1)} \sum_{\ell=0}^{K^\star} n_{0,\ell}(\mu_\ell^\star - \mu_0^\star)^2. \tag{26}
\end{aligned}
$$

Since $|D_k| = \sum_{\ell=0}^{K^\star} |D_k \cap D_\ell^\star| = \sum_{\ell=0}^{K^\star} n_{k,\ell} = n_k$,

$$
\mathbb{E}\left[\overline{Y}_{D_k}\right] = \frac{1}{n_k} \sum_{(i,j)\in D_k} \mathbb{E}\left[Y_{i,j}\right] = \frac{1}{n_k} \sum_{\ell=0}^{K^\star} \sum_{(i,j)\in D_k\cap D_\ell^\star} \mathbb{E}\left[Y_{i,j}\right] = \frac{1}{n_k} \sum_{\ell=0}^{K^\star} \mu_\ell^\star n_{k,\ell}, \tag{27}
$$

where we use for all $k \in \{1,\dots,K\}$, $D_k \subset \left(\bigcup_{\ell=0}^{K^\star} D_\ell^\star\right)$. Thus,

$$
\begin{aligned}
\sum_{(i,j)\in D_k} \left(\mathbb{E}\left[Y_{i,j}\right] - \mathbb{E}\left[\overline{Y}_{D_k}\right]\right)^2 &= \frac{1}{n_k^2} \sum_{(i,j)\in D_k} \left(n_k \mathbb{E}\left[Y_{i,j}\right] - \sum_{\ell'=0}^{K^\star} \mu_{\ell'}^\star n_{k,\ell'}\right)^2 \\
&= \frac{1}{n_k^2} \sum_{\ell=0}^{K^\star} \sum_{(i,j)\in D_k\cap D_\ell^\star} \left(n_k \mathbb{E}\left[Y_{i,j}\right] - \sum_{\ell'=0}^{K^\star} \mu_{\ell'}^\star n_{k,\ell'}\right)^2 = \frac{1}{n_k^2} \sum_{\ell=0}^{K^\star} n_{k,\ell} \left[\sum_{\ell'=0}^{K^\star} n_{k,\ell'}\left(\mu_\ell^\star - \mu_{\ell'}^\star\right)\right]^2 \\
&= \frac{1}{n_k^2} \sum_{\ell=0}^{K^\star} \sum_{\ell_1=0}^{K^\star} \sum_{\ell_2=0}^{K^\star} n_{k,\ell} n_{k,\ell_1} n_{k,\ell_2} \left(\mu_\ell^\star - \mu_{\ell_1}^\star\right)\left(\mu_\ell^\star - \mu_{\ell_2}^\star\right) \\
&= \frac{1}{n_k^2} \sum_{\ell=0}^{K^\star} \sum_{\ell_1=0}^{K^\star} n_{k,\ell} n_{k,\ell_1} \left(\mu_\ell^\star - \mu_{\ell_1}^\star\right) \sum_{\ell_2=0}^{K^\star} n_{k,\ell_2} \left(\mu_\ell^\star - \mu_{\ell_2}^\star\right) \\
&= \frac{1}{n_k} \sum_{\ell=0}^{K^\star} \sum_{\ell_1=0}^{K^\star} n_{k,\ell} n_{k,\ell_1} \left(\mu_\ell^{\star 2} - \mu_{\ell_1}^\star \mu_\ell^\star\right) - \frac{1}{n_k^2} \sum_{\ell_2=0}^{K^\star} n_{k,\ell_2} \mu_{\ell_2}^\star \underbrace{\sum_{\ell=0}^{K^\star} \sum_{\ell_1=0}^{K^\star} n_{k,\ell} n_{k,\ell_1} \left(\mu_\ell^\star - \mu_{\ell_1}^\star\right)}_{=0} \\
&= \frac{1}{n_k} \sum_{\ell=0}^{K^\star} \sum_{\ell_1=0}^{K^\star} n_{k,\ell} n_{k,\ell_1} \left(\mu_\ell^{\star 2} - \mu_{\ell_1}^\star \mu_\ell^\star\right) = \frac{1}{2n_k} \sum_{\ell=0}^{K^\star} \sum_{\ell'=0}^{K^\star} n_{k,\ell} n_{k,\ell'} \left(\mu_\ell^\star - \mu_{\ell'}^\star\right)^2.
\end{aligned}
$$

Hence,

$$\mathrm{B}_n^D(\mathbf{t}) = \frac{1}{n(n+1)} \sum_{k=1}^{K} \frac{1}{n_k} \sum_{\ell=0}^{K^\star} \sum_{\ell'=0}^{K^\star} n_{k,\ell} n_{k,\ell'} \left(\mu_\ell^\star - \mu_{\ell'}^\star \right)^2. \tag{28}$$
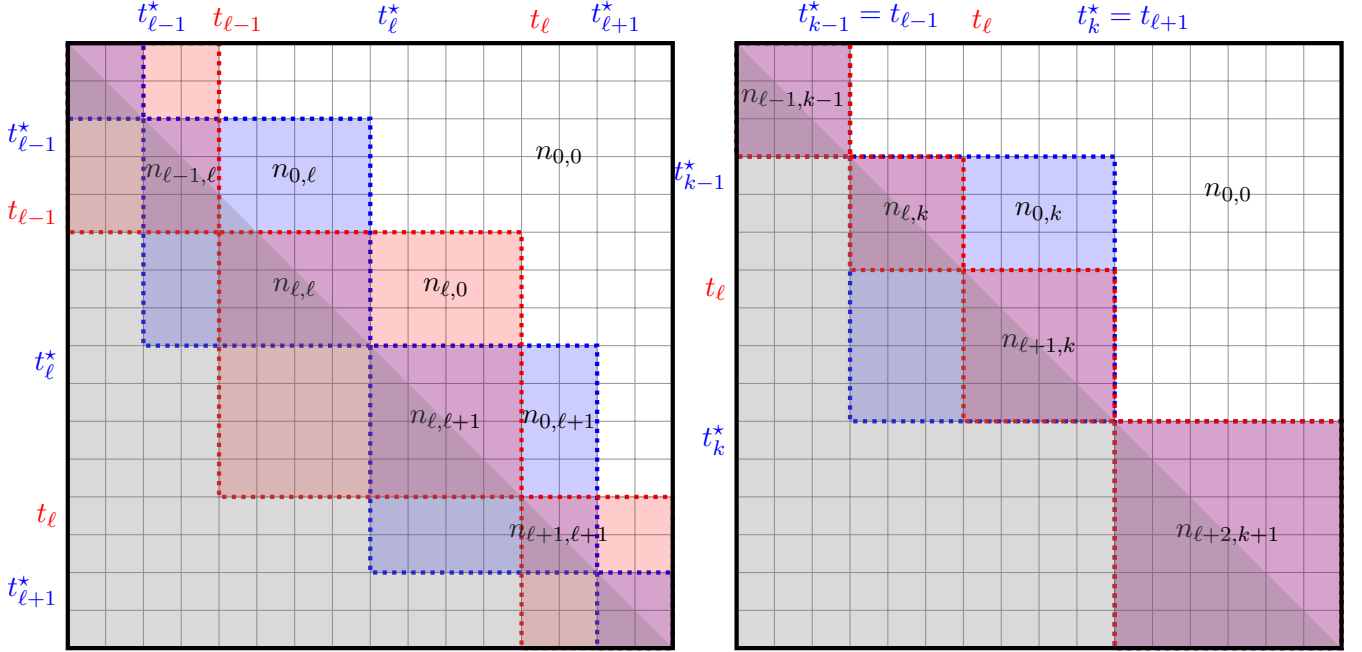

FIGURE 6.   Left: $K < K^\star$. Right: $K > K^\star$.

6.2.1. *Case $K < K^\star$ and $t \in \mathcal{A}_{n,K}^{1/n}$.* Observe that $\mathrm{B}_n(\mathbf{t}) \geq \mathrm{B}_n^D(\mathbf{t})$. Since $K < K^\star$, $t_K - t_K^\star = t_{K^\star}^\star - t_K^\star \geq n\Delta_{\boldsymbol{\tau}}^\star$. Hence, $\{k, t_k - t_k^\star \geq n\Delta_{\boldsymbol{\tau}}^\star/2\} \neq \emptyset$. Let $\ell = \min\{k, t_k - t_k^\star \geq \Delta_{\boldsymbol{\tau}}^\star/2\}$, then $\ell \geq 1$ and

$$t_{\ell-1} \leq t_\ell^\star - n\Delta_{\boldsymbol{\tau}}^\star/2 \leq t_\ell^\star + n\Delta_{\boldsymbol{\tau}}^\star/2 \leq t_\ell.$$

By definition of $\Delta_{\boldsymbol{\tau}}^\star$,

$$
\begin{aligned}
n_{\ell,\ell} &= |D_\ell \cap D_\ell^\star| \geq \min\{(t_\ell^\star - t_{\ell-1})(t_\ell^\star - t_{\ell-1} + 1)/2, (t_\ell^\star - t_{\ell-1}^\star)(t_\ell^\star - t_{\ell-1}^\star + 1)/2\} \\
&\geq (n\Delta_{\boldsymbol{\tau}}^\star)^2/8, \tag{29}
\end{aligned}
$$

and

$$n_{\ell,0} \geq \min\{(t_\ell - t_\ell^\star)(t_\ell^\star - t_{\ell-1}), (t_{\ell+1}^\star - t_\ell^\star)(t_\ell^\star - t_{\ell-1}^\star)\} \geq (n\Delta_{\boldsymbol{\tau}}^\star)^2/4. \tag{30}$$

Thus, using (29) and (30), we obtain

$$\mathrm{B}_n(\mathbf{t}) \geq \frac{1}{n(n+1)n_\ell} \left[n_{\ell,\ell} n_{\ell,0} \left(\mu_0^\star - \mu_\ell^\star \right)^2 \right] \geq \frac{\underline{\lambda}^{(0)^2}}{n(n+1)n_\ell} \frac{(n\Delta_{\boldsymbol{\tau}}^\star)^4}{32} \geq \frac{(\Delta_{\boldsymbol{\tau}}^\star)^4 \underline{\lambda}^{(0)^2}}{64},$$

since $n_\ell \leq n(n+1)/2$.

6.2.2. *Case $K > K^\star$ and $\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}$.* We have

$$B_n(\mathbf{t}) \geq B_n^0(\mathbf{t}) \geq \frac{2}{n(n+1)} n_{0,k} (\mu_k^\star - \mu_0^\star)^2 \geq \frac{2}{n(n+1)} \, n_{0,k} \, \underline{\lambda}^{(0)2}$$

for any $k \in \{0, \ldots, K^\star\}$. Since $\mathbf{t} \in \mathcal{A}_{n,K}^{\Delta_n}$, there exists $\ell \in \{1, \ldots, K-1\}$ such that for all $k \in \{0, \ldots, K^\star\}$

$$|t_k^\star - t_\ell| > \frac{n\Delta_n}{2},$$

(otherwise, it will imply that $K \leq K^\star$). Moreover, let us choose k such that $t_{k-1}^\star + n\Delta_n/2 < t_\ell < t_k^\star - n\Delta_n/2$ then

$$n_{0,k} \geq \left(t_\ell - t_{k-1}^\star\right)\left(t_k^\star - t_\ell\right) \geq \left(\frac{n\Delta_n}{2}\right)^2.$$

This leads to

$$B_n(\mathbf{t}) \geq \frac{1}{4}\underline{\lambda}^{(0)2}\Delta_n^2.$$

6.2.3. *Case $K = K^\star$ and $t \in \mathcal{A}_{n,K}^{1/n}, \|\mathbf{t} - \mathbf{t}^\star\|_\infty > n\delta$.* We have

$$B_n(t) \geq \frac{1}{n(n+1)} \frac{1}{n_\ell} n_{\ell,\ell'} n_{\ell,0}(\mu_0^\star - \mu_{\ell'}^\star)^2 \tag{31}$$

for every $\ell \in \{1, \ldots, K\}$ and every $\ell' \in \{1, \ldots, K^\star\}$. Then, we shall consider two cases: *i)* $\|\mathbf{t} - \mathbf{t}^\star\|_\infty < \frac{n\Delta_\tau^\star}{2}$ and *ii)* $\|\mathbf{t} - \mathbf{t}^\star\|_\infty \geq \frac{n\Delta_\tau^\star}{2}$.

i) $\|\mathbf{t} - \mathbf{t}^\star\|_\infty < \frac{n\Delta_\tau^\star}{2}$.

We shall assume that $t_k - t_k^\star = \|\mathbf{t} - \mathbf{t}^\star\|_\infty > 0$.

There are two possible configurations (see Figure 7). If $t_{k-1}^\star < t_{k-1} < t_k^\star < t_k$, then, by definition of $\Delta_\tau^\star$, we obtain

$$n_{k,k} = \frac{(t_k^\star - t_{k-1})(t_k^\star - t_{k-1} + 1)}{2} \geq \frac{\left(\overbrace{(t_k^\star - t_{k-1}^\star)}^{\geq n\Delta_\tau^\star} - \overbrace{(t_{k-1} - t_{k-1}^\star)}^{\leq \|\mathbf{t} - \mathbf{t}^\star\|_\infty}\right)(t_k^\star - t_{k-1} + 1)}{2} \geq \frac{(n\Delta_\tau^\star)^2}{8}. \tag{32}$$

Otherwise, if $t_{k-1} < t_{k-1}^\star < t_k^\star < t_k$, we obtain

$$n_{k,k} = \frac{(t_k^\star - t_{k-1}^\star)(t_k^\star - t_{k-1}^\star + 1)}{2} \geq \frac{(n\Delta_\tau^\star)^2}{2} \geq \frac{(n\Delta_\tau^\star)^2}{8}. \tag{33}$$

FIGURE 7. $K = K^\star$ and $\|\mathbf{t} - \mathbf{t}^\star\|_\infty < \frac{n\Delta_{\boldsymbol{\tau}}^\star}{2}$. Left: $t_{k-1}^\star < t_{k-1} < t_k^\star < t_k$. Right: $t_{k-1} < t_{k-1}^\star < t_k^\star < t_k$.

Then, by using the above decomposition of $(t_k^\star - t_{k-1})$, we obtain

$$
\begin{aligned}
n_{k,0} \;\geq\;& (t_k^\star - t_{k-1})(t_k - t_k^\star), \\
\geq\;& \left(\underbrace{(t_k^\star - t_{k-1}^\star)}_{\geq n\Delta_{\boldsymbol{\tau}}^\star} - \underbrace{(t_{k-1} - t_{k-1}^\star)}_{\leq \|\mathbf{t}-\mathbf{t}^\star\|_\infty} \right) \underbrace{(t_k - t_k^\star)}_{=\|\mathbf{t}-\mathbf{t}^\star\|_\infty} \geq \frac{\Delta_{\boldsymbol{\tau}}^\star}{2} n^2 \delta.
\end{aligned}
\tag{34}
$$

By choosing $(\ell = k, \ell' = k)$ in (31), and by using (32), (33) and (34), we obtain

$$
B_n(t) \geq \frac{1}{n(n+1)} \frac{1}{n_k} \frac{(n\Delta_{\boldsymbol{\tau}}^\star)^2}{8} \frac{\Delta_{\boldsymbol{\tau}}^\star}{2} n^2 \delta \underline{\lambda}^{(0)2} \geq \frac{(\Delta_{\boldsymbol{\tau}}^\star)^3}{32} \delta \underline{\lambda}^{(0)2}.
$$

ii) $\|\mathbf{t} - \mathbf{t}^\star\|_\infty \geq \frac{n\Delta_{\boldsymbol{\tau}}^\star}{2}$.

Since $K = K^\star$, there exists k such that $t_k^\star - t_k \geq n\frac{\Delta_{\boldsymbol{\tau}}^\star}{2}$ and $t_{k+1} - t_k^\star \geq n\frac{\Delta_{\boldsymbol{\tau}}^\star}{2}$ (otherwise, this would imply that $K > K^\star$). As above, there are two possible cases, either $t_k < t_k^\star < t_{k+1}^\star < t_{k+1}$ or $t_k < t_k^\star < t_{k+1} < t_{k+1}^\star$ (see Figure 8).

If $t_k < t_k^\star < t_{k+1}^\star < t_{k+1}$, we obtain, by definition of $\Delta_{\boldsymbol{\tau}}^\star$,

$$
n_{k+1,k+1} = \frac{(t_{k+1}^\star - t_k^\star)(t_{k+1}^\star - t_k^\star + 1)}{2} \geq \frac{(n\Delta_{\boldsymbol{\tau}}^\star)^2}{2},
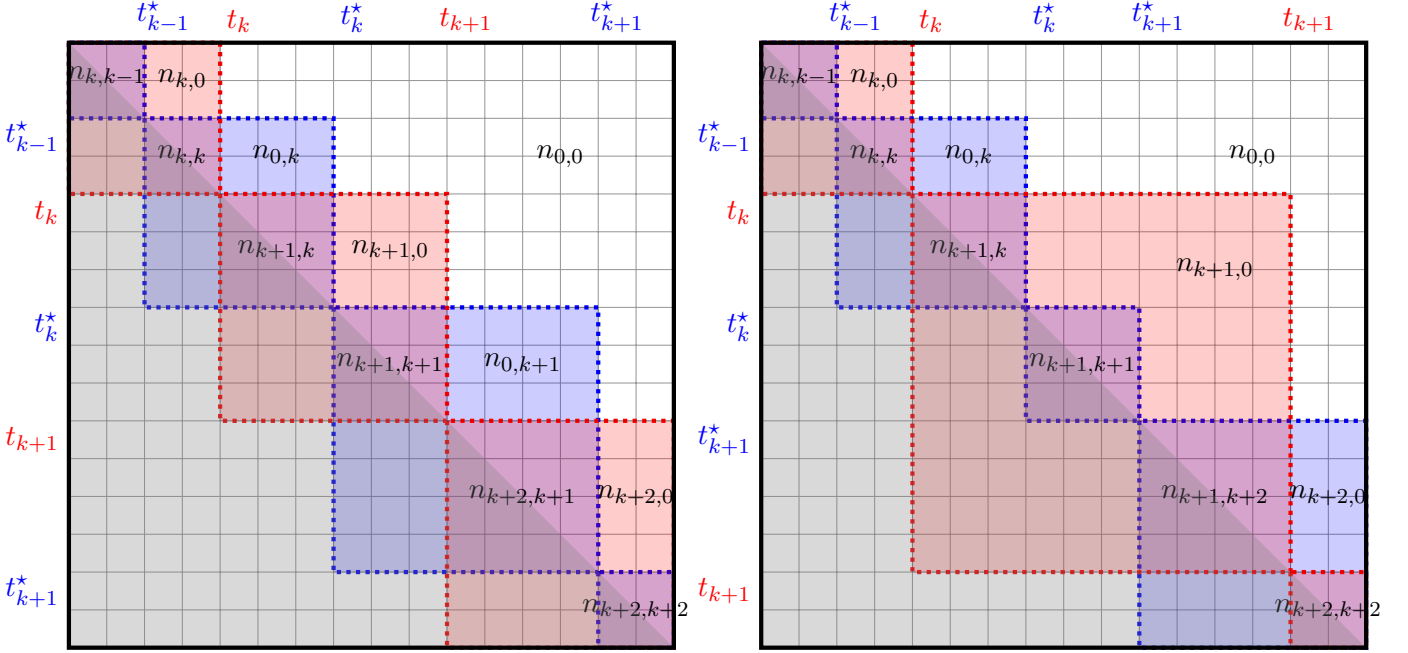\tag{35}
$$

FIGURE 8. $K = K^\star$ and $\|\mathbf{t} - \mathbf{t}^\star\|_\infty \geq \frac{n\Delta_\tau^\star}{2}$. Left: $t_k < t_k^\star < t_{k+1}^\star < t_{k+1}$.
Right: $t_k < t_k^\star < t_{k+1} < t_{k+1}^\star$.

and

$$n_{k+1,0} \geq (t_{k+1}^\star - t_k^\star)(t_k^\star - t_k) \geq (n\Delta_\tau^\star)\frac{(n\Delta_\tau^\star)}{2}. \tag{36}$$

If $t_k < t_k^\star < t_{k+1} < t_{k+1}^\star$, we obtain

$$n_{k+1,k+1} = \frac{(t_{k+1} - t_k^\star)(t_{k+1} - t_k^\star + 1)}{2} \geq \frac{1}{2}\left(\frac{n\Delta_\tau^\star}{2}\right)^2, \tag{37}$$

and

$$n_{k+1,0} \geq (t_{k+1} - t_k^\star)(t_k^\star - t_k) \geq \left(\frac{(n\Delta_\tau^\star)}{2}\right)^2. \tag{38}$$

By choosing $(\ell = \ell' = k + 1)$ in (31), and by using (35), (36), (37) and (38), we obtain

$$B_n(t) \geq \frac{(\Delta_\tau^\star)^4}{32}\underline{\lambda}^{(0)^2}.$$

6.3. Deviation inequalities.

Lemma 2. *For all $\alpha > 0$,*

$$\mathbb{P}\left(-\min_{\mathbf{t}\in\mathcal{A}_{n,K}^{1/n}} V_n(\mathbf{t}) \geq \alpha\right) \leq n(n+1)e^{-\frac{n(n+1)\alpha}{16K\beta}} + 2e^{-\frac{|G_{01}|\alpha}{8\beta}},$$

where V_n is defined by (20) and (22) and $\mathcal{A}_{n,K}^{1/n}$ is defined in (6) with $\Delta_n = 1/n$. Moreover, if $\alpha = \alpha_n$ is such that $\alpha_n n^2 / \log(n) \to \infty$ and $\alpha_n |G_{0,1}| \to \infty$, as n tends to infinity, then

$$\mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}} V_n(\mathbf{t}) \geq \alpha_n\right) \to 0, \ \text{as } n \to +\infty.$$

The proof is given in the supplementary material.

Lemma 3. *Let W_n be defined by (20) and (23), then there exists $C_1 > 0$ such that for all $\alpha > 0$:*

$$\mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}} W_n(\mathbf{t}) \geq \alpha\right) \ \leq \ C_1 n^{4K_{\max}} \exp\left[-\frac{\alpha^2 n(n+1)}{128\,\beta\,(K+1)^2\,(K^\star+1)^2\,\overline{\lambda}^2}\right],$$

where $\overline{\lambda} = \sup\limits_{k \neq \ell} |\mu_k^\star - \mu_\ell^\star|$ and $\mathcal{A}_{n,K}^{1/n}$ is defined in (6) with $\Delta_n = 1/n$. Moreover, if $\alpha = \alpha_n$ is such that $\alpha_n^2 n^2 / \log(n) \to \infty$, as n tends to infinity, then

$$\mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}} W_n(\mathbf{t}) \geq \alpha_n\right) \to 0, \ \text{as } n \to \infty.$$

The proof is given in the supplementary material.

Lemma 4. *For all $\alpha > 0$ and $\gamma > 0$,*

$$\mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}} Z_n(\mathbf{t}) \geq \alpha\right) \leq 2e^{-\frac{|G_{01}|\gamma^2}{4\beta}} + 2C_1 n^{4K_{\max}} e^{-\frac{\alpha^2 n(n+1)}{512\gamma^2\beta}} + 2e^{-\frac{|G_{01}|\alpha^2 n^2}{32\overline{\lambda}^2\beta}},$$

where Z_n is defined by (24), $\mathcal{A}_{n,K}^{1/n}$ is defined in (6) with $\Delta_n = 1/n$ and $\overline{\lambda} = \sup\limits_{k \neq \ell} |\mu_k^\star - \mu_\ell^\star|$. Moreover, if $\alpha = \alpha_n$ is such that $\alpha_n^2 n^2 / \log(n) \to \infty$ and $\alpha_n^2 n^2 |G_{0,1}| \to \infty$, as n tends to infinity, then

$$\mathbb{P}\left(-\min_{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}} Z_n(\mathbf{t}) \geq \alpha_n\right) \to 0, \ \text{as } n \to \infty.$$

The proof is given in the supplementary material.

References

M. Basseville and I. Nikiforov. *The detection of abrupt changes.* Information and System Sciences Series. Prentice-Hall., 1993.

R. Bellman. On the approximation of curves by line segments using dynamic programming. *Communications of the ACM*, 4(6):284, 1961.

L. Boysen, A. Kempe, V. Liebscher, A. Munk, and O. Wittich. Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.*, 37(1):157–183, 02 2009. doi: 10.1214/07-AOS558. URL `http://dx.doi.org/10.1214/07-AOS558`.

B. Brodsky and B. Darkhovsky. *Non-parametric statistical diagnosis: problems and methods.* Kluwer Academic Publishers, 2000.

A. Dehman, C. Ambroise, and P. Neuvial. Performance of a blockwise approach in variable selection using linkage disequilibrium information. *BMC Bioinformatics*, 16(148), 2015.

J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.

E. Ioanna Delatola, E. Lebarbier, T. Mary-Huard, F. Radvanyi, S. Robin, and J. Wong. Seg-Corr: a statistical procedure for the detection of genomic regions of correlated expression. *ArXiv e-prints*, 2015.

M. Lavielle. Using penalized contrasts for the change-point problem. *Signal Processing*, 85 (8):1501–1510, 2005.

M. Lavielle and E. Moulines. Least-squares estimation of an unknown number of shifts in a time series. *Journal of time series analysis*, 21(1):33–59, 2000.

E. Lebarbier. Detecting multiple change-points in the mean of a Gaussian process by model selection. *Signal Processing*, 85:717–736, 2005.

C. Lévy-Leduc, M. Delattre, T. Mary-Huard, and S. Robin. Two-dimensional segmentation for analyzing HiC data. *Bioinformatics*, 30(17):386–392, 2014.

P. Massart. A non asymptotic theory for model selection. In *Proceedings of Mathematical Foundations of Statistical Learning, Stockholm*, pages 309–324, 2004.

A. Tartakovsky, I. Nikiforov, and M. Basseville. *Sequential Analysis: Hypothesis Testing and Changepoint Detection.* CRC Press, Taylor & Francis Group, 2014.

Y. Yao and S. T. Au. Least-squares estimation of a step function. *Sankhya: The Indian Journal of Statistics, Series A*, 51(3):370–381, 1989.

Proof of Equation (14). The goal of this section is to prove that J_n defined in (13) can be rewritten as in (14). By definition of Q_n^K given in (10),

$$J_n(\mathbf{t}) = J_n^D(\mathbf{t}) + J_n^0(\mathbf{t}), \tag{39}$$

where

$$J_n^D(\mathbf{t}) = \frac{2}{n(n+1)} \left(\left[\sum_{k=1}^{K} \sum_{(i,j) \in D_k} \left(Y_{i,j} - \overline{Y}_{D_k} \right)^2 \right] - \left[\sum_{k=1}^{K^\star} \sum_{(i,j) \in D_k^\star} \left(Y_{i,j} - \overline{Y}_{D_k^\star} \right)^2 \right] \right),$$

and

$$J_n^0(\mathbf{t}) = \frac{2}{n(n+1)} \left(\sum_{(i,j) \in E_0} \left(Y_{i,j} - \overline{Y}_{G_{01}} \right)^2 - \sum_{(i,j) \in E_0^\star} \left(Y_{i,j} - \overline{Y}_{G_{01}} \right)^2 \right).$$

Using (1) and $\overline{Y}_{D_k} = \mathbb{E}\left[\overline{Y}_{D_k} \right] + |D_k|^{-1} \sum_{(i',j') \in D_k} \varepsilon_{i',j'}$, where $|A|$ denotes the cardinality of the set A, we obtain

$$\sum_{(i,j) \in D_k} \left(Y_{i,j} - \overline{Y}_{D_k} \right)^2 = \sum_{(i,j) \in D_k} \left[Y_{i,j}^2 - 2 Y_{i,j} \overline{Y}_{D_k} + \overline{Y}_{D_k}^2 \right]$$

$$= \sum_{(i,j) \in D_k} \left[\mathbb{E}\left[Y_{i,j} \right]^2 + 2 \mathbb{E}\left[Y_{i,j} \right] \varepsilon_{i,j} + \varepsilon_{i,j}^2 \right]$$

$$-2 \sum_{(i,j) \in D_k} \left[\mathbb{E}\left[Y_{i,j} \right] \mathbb{E}\left[\overline{Y}_{D_k} \right] + \varepsilon_{i,j} \mathbb{E}\left[\overline{Y}_{D_k} \right] + \mathbb{E}\left[Y_{i,j} \right] \frac{1}{|D_k|} \left(\sum_{(i',j') \in D_k} \varepsilon_{i',j'} \right) + \varepsilon_{i,j} \frac{1}{|D_k|} \left(\sum_{(i',j') \in D_k} \varepsilon_{i',j'} \right) \right]$$

$$+ \sum_{(i,j) \in D_k} \left[\mathbb{E}\left[\overline{Y}_{D_k} \right]^2 + 2 \mathbb{E}\left[\overline{Y}_{D_k} \right] \frac{1}{|D_k|} \left(\sum_{(i',j') \in D_k} \varepsilon_{i',j'} \right) + \frac{1}{|D_k|^2} \left(\sum_{(i',j') \in D_k} \varepsilon_{i',j'} \right)^2 \right].$$

By gathering the deterministic terms and the terms linked to the noise, we obtain

$$\sum_{(i,j) \in D_k} \left(Y_{i,j} - \overline{Y}_{D_k} \right)^2 = \sum_{(i,j) \in D_k} \left(\mathbb{E}\left[Y_{i,j} \right] - \mathbb{E}\left[\overline{Y}_{D_k} \right] \right)^2 - 2 \frac{1}{|D_k|} \left(\sum_{(i',j') \in D_k} \varepsilon_{i',j'} \right) \sum_{(i,j) \in D_k} \mathbb{E}\left[Y_{i,j} \right]$$

$$- \frac{1}{|D_k|} \left(\sum_{(i',j') \in D_k} \varepsilon_{i',j'} \right)^2 + \sum_{(i,j) \in D_k} \left(2 \mathbb{E}\left[Y_{i,j} \right] \varepsilon_{i,j} + \varepsilon_{i,j}^2 \right). \tag{40}$$

Thus, for J_n^D defined in (39), we obtain

$$J_n^D(\mathbf{t}) = \frac{2}{n(n+1)} \left[\sum_{k=1}^{K} \sum_{(i,j)\in D_k} \left(\mathbb{E}\left[Y_{i,j}\right] - \mathbb{E}\left[\overline{Y}_{D_k}\right]\right)^2 - \underbrace{\sum_{k=1}^{K^\star} \sum_{(i,j)\in D_k^\star} \left(\mathbb{E}\left[Y_{i,j}\right] - \mathbb{E}\left[\overline{Y}_{D_k^\star}\right]\right)^2}_{=0} \right]$$

$$- \frac{4}{n(n+1)} \left[\sum_{k=1}^{K} \frac{1}{|D_k|} \left(\sum_{(i',j')\in D_k} \varepsilon_{i',j'} \right) \sum_{(i,j)\in D_k} \mathbb{E}\left[Y_{i,j}\right] - \sum_{k=1}^{K^\star} \frac{1}{|D_k^\star|} \left(\sum_{(i',j')\in D_k^\star} \varepsilon_{i',j'} \right) \sum_{(i,j)\in D_k^\star} \mathbb{E}\left[Y_{i,j}\right] \right]$$

$$- \frac{2}{n(n+1)} \left[\sum_{k=1}^{K} \frac{1}{|D_k|} \left(\sum_{(i',j')\in D_k} \varepsilon_{i',j'} \right)^2 - \sum_{k=1}^{K^\star} \frac{1}{|D_k^\star|} \left(\sum_{(i',j')\in D_k^\star} \varepsilon_{i',j'} \right)^2 \right]$$

$$+ \frac{2}{n(n+1)} \left(\sum_{k=1}^{K} \sum_{(i,j)\in D_k} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2 \right) - \sum_{k=1}^{K^\star} \sum_{(i,j)\in D_k^\star} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2 \right) \right)$$

$$= \mathrm{B}_n^D(\mathbf{t}) + W_n^D(\mathbf{t}) + V_n^D(\mathbf{t})$$

$$+ \frac{2}{n(n+1)} \left(\sum_{k=1}^{K} \sum_{(i,j)\in D_k} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2 \right) - \sum_{k=1}^{K^\star} \sum_{(i,j)\in D_k^\star} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2 \right) \right), \tag{41}$$

since $\mathbb{E}[Y_{i,j}] = \mathbb{E}[\overline{Y}_{D_k^\star}]$, for all $(i,j) \in D_k^\star$. Using (8), we obtain

$$J_n^0(\mathbf{t}) = \frac{2}{n(n+1)} \left(\sum_{(i,j)\in G_{00}} \left(Y_{i,j} - \overline{Y}_{G_{01}} \right)^2 - \sum_{(i,j)\in G_{00}^\star} \left(Y_{i,j} - \overline{Y}_{G_{01}} \right)^2 \right).$$

Using (1) and $\overline{Y}_{G_{01}} = \mathbb{E}\left[\overline{Y}_{G_{01}}\right] + |G_{01}|^{-1} \sum_{(i',j')\in G_{01}} \varepsilon_{i',j'}$, we obtain

$$\sum_{(i,j)\in G_{00}} \left(Y_{i,j} - \overline{Y}_{G_{01}} \right)^2 = \sum_{(i,j)\in G_{00}} \left[Y_{i,j}^2 - 2Y_{i,j}\overline{Y}_{G_{01}} + \overline{Y}_{G_{01}}^2 \right]$$

$$= \sum_{(i,j)\in G_{00}} \left[\mathbb{E}\left[Y_{i,j}\right]^2 + 2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2 \right]$$

$$- 2 \sum_{(i,j)\in G_{00}} \left\{ \mathbb{E}\left[Y_{i,j}\right]\mathbb{E}\left[\overline{Y}_{G_{01}}\right] + \varepsilon_{i,j}\mathbb{E}\left[\overline{Y}_{G_{01}}\right] + \mathbb{E}\left[Y_{i,j}\right] \frac{1}{|G_{01}|} \left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'} \right) \right.$$

$$\left. + \varepsilon_{i,j} \frac{1}{|G_{01}|} \left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'} \right) \right\}$$

$$+ \sum_{(i,j)\in G_{00}} \left\{ \mathbb{E}\left[\overline{Y}_{G_{01}}\right]^2 + 2\mathbb{E}\left[\overline{Y}_{G_{01}}\right] \frac{1}{|G_{01}|} \left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'} \right) + \frac{1}{|G_{01}|^2} \left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'} \right)^2 \right\}.$$

By gathering the deterministic terms and the terms linked to the noise as in (40), we obtain

$$
\sum_{(i,j)\in G_{00}} \left(Y_{i,j} - \overline{Y}_{G_{01}}\right)^2 = \sum_{(i,j)\in G_{00}} \left(\mathbb{E}\left[Y_{i,j}\right] - \mathbb{E}\left[\overline{Y}_{G_{01}}\right]\right)^2
$$

$$
-2\mu_0^\star \sum_{(i,j)\in G_{00}} \varepsilon_{i,j} - 2\frac{|G_{00}|}{|G_{01}|}\left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'}\right)\left[\frac{1}{|G_{00}|}\sum_{(i,j)\in G_{00}}\mathbb{E}(Y_{i,j}) - \mu_0^\star\right]
$$

$$
+ \frac{1}{|G_{01}|}\left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'}\right)\left[\frac{|G_{00}|}{|G_{01}|}\left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'}\right) - 2\left(\sum_{(i,j)\in G_{00}} \varepsilon_{i,j}\right)\right]
$$

$$
+ \sum_{(i,j)\in G_{00}} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right),
$$

where $\mu_0^\star$ is defined in (2). Thus, we obtain

$$
J_n^0(\mathbf{t}) = \frac{2}{n(n+1)}\left(\sum_{(i,j)\in G_{00}}\left(\mathbb{E}\left[Y_{i,j}\right] - \mathbb{E}\left[\overline{Y}_{G_{01}}\right]\right)^2 - \sum_{(i,j)\in G_{00}^\star}\left(\mathbb{E}\left[Y_{i,j}\right] - \mathbb{E}\left[\overline{Y}_{G_{01}}\right]\right)^2\right)
$$

$$
-\frac{4}{n(n+1)}\mu_0^\star\left(\sum_{(i,j)\in G_{00}} \varepsilon_{i,j} - \sum_{(i,j)\in G_{00}^\star} \varepsilon_{i,j}\right)
$$

$$
+\frac{2}{n(n+1)}\frac{1}{|G_{01}|^2}\left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'}\right)^2 (|G_{00}| - |G_{00}^\star|)
$$

$$
-\frac{4}{n(n+1)}\frac{1}{|G_{01}|}\left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'}\right)\left[\sum_{(i,j)\in G_{00}}\mathbb{E}(Y_{i,j}) - \sum_{(i,j)\in G_{00}^\star}\mathbb{E}(Y_{i,j}) - \mu_0^\star(|G_{00}| - |G_{00}^\star|)\right]
$$

$$
-\frac{4}{n(n+1)}\frac{1}{|G_{01}|}\left(\sum_{(i',j')\in G_{01}} \varepsilon_{i',j'}\right)\left(\sum_{(i,j)\in G_{00}} \varepsilon_{i,j} - \sum_{(i,j)\in G_{00}^\star} \varepsilon_{i,j}\right)
$$

$$
+\frac{2}{n(n+1)}\left(\sum_{(i,j)\in G_{00}}\left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) - \sum_{(i,j)\in G_{00}^\star}\left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right)\right)
$$

$$
= \ \mathrm{B}_n^0(\mathbf{t}) + W_n^0(\mathbf{t}) + V_n^0(\mathbf{t}) + Z_n(\mathbf{t})
$$

$$
+\frac{2}{n(n+1)}\left(\sum_{(i,j)\in G_{00}}\left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) - \sum_{(i,j)\in G_{00}^\star}\left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right)\right), \tag{42}
$$

since $\mathbb{E}[Y_{i,j}] = \mu_0^\star = \mathbb{E}\left[\overline{Y}_{G_{01}}\right]$, for all $(i,j) \in G_{00}^\star$. Then, from (39), (41) and (42), we obtain

$$J_n(\mathbf{t}) = \mathrm{B}_n^D(\mathbf{t}) + W_n^D(\mathbf{t}) + V_n^D(\mathbf{t}) + \mathrm{B}_n^0(\mathbf{t}) + W_n^0(\mathbf{t}) + V_n^0(\mathbf{t}) + Z_n(\mathbf{t})$$

$$+ \quad \frac{2}{n(n+1)} \left(\sum_{k=1}^{K} \sum_{(i,j)\in D_k} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) - \sum_{k=1}^{K^\star} \sum_{(i,j)\in D_k^\star} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) \right)$$

$$+ \quad \frac{2}{n(n+1)} \left(\sum_{(i,j)\in G_{00}} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) - \sum_{(i,j)\in G_{00}^\star} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) \right).$$

Note that

$$\sum_{k=1}^{K} \sum_{(i,j)\in D_k} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) + \sum_{(i,j)\in G_{00}} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) + \sum_{(i,j)\in G_{01}} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right)$$

$$= \sum_{k=1}^{K^\star} \sum_{(i,j)\in D_k^\star} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) + \sum_{(i,j)\in G_{00}^\star} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right) + \sum_{(i,j)\in G_{01}} \left(2\mathbb{E}\left[Y_{i,j}\right]\varepsilon_{i,j} + \varepsilon_{i,j}^2\right),$$

since it amounts to summing up over all the possible indices i and j. This concludes the proof.

Proof of Lemma 2. By (20) and (22),

$$V_n(\mathbf{t}) = \left\{ \frac{2}{n(n+1)} \left[\sum_{k=1}^{K^\star} \frac{\left(\sum_{(i,j)\in D_k^\star} \varepsilon_{i,j}\right)^2}{|D_k^\star|} - \sum_{k'=1}^{K} \frac{\left(\sum_{(i',j')\in D_{k'}} \varepsilon_{i',j'}\right)^2}{|D_{k'}|} \right] \right\}$$

$$+ \left\{ \frac{2}{n(n+1)} \frac{1}{|G_{01}|^2} \left(\sum_{(i,j)\in G_{01}} \varepsilon_{i,j} \right)^2 \left(|G_{00}| - |G_{00}^\star|\right) \right\}$$

$$= V_n^D(\mathbf{t}) + V_n^0(\mathbf{t}).$$

Hence,

$$\mathbb{P}\left(-\min_{\mathbf{t}\in\mathcal{A}_{n,K}^{1/n}} V_n(\mathbf{t}) \geq \alpha\right) \leq \mathbb{P}\left(-\min_{\mathbf{t}\in\mathcal{A}_{n,K}^{1/n}} V_n^D(\mathbf{t}) \geq \frac{\alpha}{2}\right) + \mathbb{P}\left(-\min_{\mathbf{t}\in\mathcal{A}_{n,K}^{1/n}} V_n^0(\mathbf{t}) \geq \frac{\alpha}{2}\right)$$

$$= \mathbb{P}\left(\max_{\mathbf{t}\in\mathcal{A}_{n,K}^{1/n}} -V_n^D(\mathbf{t}) \geq \frac{\alpha}{2}\right) + \mathbb{P}\left(\max_{\mathbf{t}\in\mathcal{A}_{n,K}^{1/n}} -V_n^0(\mathbf{t}) \geq \frac{\alpha}{2}\right). \quad (43)$$

Let us now derive an upper bound for each of these two probabilities. Let us first address the first term in the rhs of (43). Note that

$$
V_n^D(\mathbf{t}) \geq -\frac{2}{n(n+1)} \left[\sum_{k=1}^K \frac{\left(\sum_{(i,j)\in D_k} \varepsilon_{i,j}\right)^2}{|D_k|} \right] \geq -\frac{2}{n(n+1)} \left[\sum_{k=1}^K \max_{A \in \mathcal{I}_D} \frac{\left(\sum_{(i,j)\in A} \varepsilon_{i,j}\right)^2}{|A|} \right]
$$

$$
\geq -\frac{2K}{n(n+1)} \left[\max_{A \in \mathcal{I}_D} \frac{\left(\sum_{(i,j)\in A} \varepsilon_{i,j}\right)^2}{|A|} \right],
$$

where $\mathcal{I}_D$ is the set of all possible diagonal blocks:

$$
\mathcal{I}_D = \left\{ \left\{ (i,j) \in \{1,\ldots,n\}^2 \,\middle|\, t \leq i \leq j \leq t' \right\} \,\middle|\, 1 \leq t < t' \leq n \right\}. \tag{44}
$$

Thus,

$$
\mathbb{P}\left(\max_{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}} -V_n^D(\mathbf{t}) \geq \frac{\alpha}{2} \right) \leq \mathbb{P}\left(\frac{2K}{n(n+1)} \left[\max_{A \in \mathcal{I}_D} \frac{\left(\sum_{(i,j)\in A} \varepsilon_{i,j}\right)^2}{|A|} \right] \geq \frac{\alpha}{2} \right).
$$

Moreover, from Lemma 5 we obtain that for all positive u,

$$
\mathbb{P}\left(\max_{A \in \mathcal{I}_D} \frac{\left(\sum_{(i,j)\in A} \varepsilon_{i,j}\right)^2}{|A|} \geq u \right) \leq \sum_{A \in \mathcal{I}_D} \mathbb{P}\left(\left| \sum_{(i,j)\in A} \varepsilon_{i,j} \right| \geq \sqrt{u|A|} \right) \leq 2 \sum_{A \in \mathcal{I}_D} e^{-\frac{u}{4\beta}}
$$

$$
\leq n(n+1) e^{-\frac{u}{4\beta}},
$$

where we use $|\mathcal{I}_D| \leq n(n+1)/2$. Setting $u = \frac{n(n+1)\alpha}{4K}$ in the previous inequality yields

$$
\mathbb{P}\left(\max_{\mathbf{t} \in \mathcal{A}_{n,K}^{1/n}} -V_n^D(\mathbf{t}) \geq \frac{\alpha}{2} \right) \leq n(n+1) e^{-\frac{n(n+1)\alpha}{16K\beta}}. \tag{45}
$$

Let us now address the second term in the rhs of (43). Note that

$$
V_n^0(\mathbf{t}) = \frac{2}{n(n+1)} \frac{1}{|G_{01}|^2} \left(\sum_{(i,j)\in G_{01}} \varepsilon_{i,j} \right)^2 (|G_{00}| - |G_{00}^\star|) \geq -\frac{2|G_{00}^\star|}{n(n+1)} \frac{1}{|G_{01}|^2} \left(\sum_{(i,j)\in G_{01}} \varepsilon_{i,j} \right)^2
$$

$$
\geq -\left(\frac{\sum_{(i,j)\in G_{01}} \varepsilon_{i,j}}{|G_{01}|} \right)^2,
$$

which leads to

$$
\mathbb{P}\left(\max_{\mathbf{t} \in \mathcal{A}_{n,K}} -V_n^0(\mathbf{t}) \geq \frac{\alpha}{2} \right) \leq \mathbb{P}\left(\left| \frac{\sum_{(i,j)\in G_{01}} \varepsilon_{i,j}}{|G_{01}|} \right| \geq \sqrt{\frac{\alpha}{2}} \right) \leq 2 e^{-\frac{|G_{01}|\alpha}{8\beta}}, \tag{46}
$$

where the last inequality comes from Lemma 5. The conclusion follows from (45) and (46).

Proof of Lemma 3. Using (27) W_n can be rewritten as follows:

$$W_n(\mathbf{t}) = \frac{4}{n(n+1)} \left[\sum_{k'=0}^{K^\star} \left(\sum_{(i',j')\in D_{k'}^\star} \varepsilon_{i',j'} \right) \mu_{k'}^\star - \sum_{k=0}^{K} \left\{ \left(\sum_{(i,j)\in D_k} \varepsilon_{i,j} \right) \sum_{\ell=0}^{K^\star} \frac{n_{k,\ell}}{n_k} \mu_\ell^\star \right\} \right]$$

$$+ \frac{4}{n(n+1)} \left(\sum_{(i,j)\in D_0} \varepsilon_{i,j} \right) \sum_{\ell=0}^{K^\star} \frac{n_{0,\ell}}{n_0} (\mu_\ell^\star - \mu_0^\star), \tag{47}$$

where we used that $\mu_0^\star = \sum_{\ell=0}^{K^\star} \frac{n_{0,\ell}}{n_0} \mu_0^\star$. With the notation

$$e_{k,\ell} = \sum_{(i,j)\in D_k^\star \cap D_\ell} \varepsilon_{i,j},$$

the first term in the rhs of the previous equation can be rewritten as follows:

$$\frac{4}{n(n+1)} \left[\sum_{k'=0}^{K^\star} \left(\sum_{\ell'=0}^{K} \sum_{(i',j')\in D_{k'}^\star \cap D_{\ell'}} \varepsilon_{i',j'} \right) \mu_{k'}^\star - \sum_{k=0}^{K} \left[\left(\sum_{k_1=0}^{K^\star} \sum_{(i,j)\in D_{k_1}^\star \cap D_k} \varepsilon_{i,j} \right) \sum_{\ell=0}^{K^\star} \frac{n_{k,\ell}}{n_k} \mu_\ell^\star \right] \right]$$

$$= \frac{4}{n(n+1)} \left[\sum_{k'=0}^{K^\star} \sum_{\ell'=0}^{K} e_{k',\ell'} \sum_{\ell_1=0}^{K^\star} \frac{n_{\ell',\ell_1}}{n_{\ell'}} \mu_{k'}^\star - \sum_{k=0}^{K} \sum_{k_1=0}^{K^\star} e_{k_1,k} \sum_{\ell=0}^{K^\star} \frac{n_{k,\ell}}{n_k} \mu_\ell^\star \right] \tag{48}$$

where we used that $\sum_{\ell_1=0}^{K^\star} n_{\ell',\ell_1}/n_{\ell'} \mu_{k'}^\star = \mu_{k'}^\star$. We deduce from (47) and (48) that

$$W_n(\mathbf{t}) = \frac{4}{n(n+1)} \sum_{k=0}^{K} \sum_{\ell'=0}^{K^\star} \sum_{\ell=0}^{K^\star} e_{\ell',k} \frac{n_{k,\ell}}{n_k} (\mu_{\ell'}^\star - \mu_\ell^\star) + \frac{4}{n(n+1)} \left(\sum_{(i,j)\in D_0} \varepsilon_{i,j} \right) \sum_{\ell=0}^{K^\star} \frac{n_{0,\ell}}{n_0} (\mu_\ell^\star - \mu_0^\star).$$

Thus,

$$|W_n(\mathbf{t})| \leq \frac{8}{n(n+1)} \sum_{k=0}^{K} \sum_{\ell'=0}^{K^\star} \sum_{\ell=0}^{K^\star} \left[\max_{A\in\mathcal{I}_R^{K_{\max}}\cup\mathcal{I}_D} \left| \sum_{(i',j')\in A} \varepsilon_{i',j'} \right| \right] \frac{n_{k,\ell}}{n_k} \overline{\lambda}$$

$$\leq \frac{8}{n(n+1)} \overline{\lambda} \left[\max_{A\in\mathcal{I}_R^{K_{\max}}\cup\mathcal{I}_D} \left| \sum_{(i',j')\in A} \varepsilon_{i',j'} \right| \right] \sum_{k=0}^{K} \sum_{\ell'=0}^{K^\star} \frac{n_k}{n_k}$$

$$\leq \frac{8\,(K+1)\,(K^\star+1)}{n(n+1)} \overline{\lambda} \left[\max_{A\in\mathcal{I}_R^{K_{\max}}\cup\mathcal{I}_D} \left| \sum_{(i',j')\in A} \varepsilon_{i',j'} \right| \right],$$

where $\mathcal{I}_D$ is defined in (44) and

$$\mathcal{I}_R^{K_{\max}} = \left\{ \left. \bigcup_{k=1}^{K_{\max}} \left\{ (i,j) \in \{1,\ldots,n\}^2 \,\middle|\, i_k^{(1)} \leq i < i_k^{(2)} \text{ and } j_k^{(1)} \leq j < j_k^{(2)} \right\} \right| \right.$$

$$\left. \forall k,\ 1 \leq i_k^{(1)} \leq i_k^{(2)} \leq j_k^{(1)} \leq j_k^{(2)} \leq n+1 \right\}. \tag{49}$$

Using Lemma 6, we obtain that

$$\mathbb{P}\left(-\min_{\mathbf{t}\in\mathcal{A}_{n,K}^{1/n}} W_n(\mathbf{t}) \geq \alpha\right) \leq \mathbb{P}\left(\frac{8\,(K+1)\,(K^\star+1)}{n(n+1)}\overline{\lambda}\left[\max_{A\in\mathcal{I}_R^{K\max}\cup\mathcal{I}_D}\left|\sum_{(i',j')\in A}\varepsilon_{i',j'}\right|\right] > \alpha\right)$$

$$\leq \left|\mathcal{I}_R^{K\max}\cup\mathcal{I}_D\right|\exp\left[-\frac{\alpha^2 n(n+1)}{128\,\beta\,(K+1)^2\,(K^\star+1)^2\,\overline{\lambda}^2}\right],$$

which concludes the proof by using that $\left|\mathcal{I}_R^{K\max}\cup\mathcal{I}_D\right| \leq \left|\mathcal{I}_R^{K\max}\right| + |\mathcal{I}_D| \leq 2\left|\mathcal{I}_R^{K\max}\right|$ and

$$|\mathcal{I}_R^{K\max}| \leq \binom{n-1}{K-1}^4 \leq C_1 n^{4K\max}, \tag{50}$$

for some positive constant C_1.

Proof of Lemma 4. Using (25), we obtain

$$\sum_{(i,j)\in G_{00}}(\mathbb{E}\,[Y_{i,j}] - \mu_0^\star) = \sum_{\ell=0}^{K^\star}\sum_{(i,j)\in G_{00}\cap D_\ell^\star}(\mathbb{E}\,[Y_{i,j}] - \mu_0^\star) = \sum_{\ell=0}^{K^\star}n_{0,\ell}(\mu_\ell^\star - \mu_0^\star) = \sum_{\ell=1}^{K^\star}n_{0,\ell}(\mu_\ell^\star - \mu_0^\star).$$

By (24), we thus obtain that

$$Z_n(\mathbf{t}) = \frac{4}{n(n+1)}\frac{1}{|G_{01}|}\left(\sum_{(i,j)\in G_{01}}\varepsilon_{i,j}\right)\left[\left(\sum_{(i,j)\in G_{00}^\star}\varepsilon_{i,j} - \sum_{(i,j)\in G_{00}}\varepsilon_{i,j}\right) - \sum_{\ell=1}^{K^\star}n_{0,\ell}(\mu_\ell^\star - \mu_0^\star)\right].$$

This gives the following upper bound for Z_n:

$$|Z_n(\mathbf{t})| \leq \frac{8}{n(n+1)}\left[\frac{\left|\sum_{(i,j)\in G_{01}}\varepsilon_{i,j}\right|}{|G_{01}|}\right] \times \left[\max_{A\in\mathcal{I}_R^{K\max}}\left|\sum_{(i,j)\in A}\varepsilon_{i,j}\right|\right] + \frac{4}{n(n+1)}\left[\frac{\left|\sum_{(i,j)\in G_{01}}\varepsilon_{i,j}\right|}{|G_{01}|}\right]n_0\overline{\lambda},$$

where $\mathcal{I}_R^{K\max}$ is defined in (49) and where we used:

$$\sum_{(i,j)\in G_{00}^\star}\varepsilon_{i,j} - \sum_{(i,j)\in G_{00}}\varepsilon_{i,j} = \sum_{(i,j)\in G_{00}^\star\cap G_{00}{}^C}\varepsilon_{i,j} - \sum_{(i,j)\in G_{00}\cap G_{00}^\star{}^C}\varepsilon_{i,j},$$

A^C denoting the complement of the set A. Thus,

$$\mathbb{P}\left(-\min_{\mathbf{t}\in\mathcal{A}_{n,K}} Z_n(\mathbf{t}) \geq \alpha\right) \leq \mathbb{P}\left(\frac{8}{n(n+1)}\left[\frac{\left|\sum_{(i,j)\in G_{01}}\varepsilon_{i,j}\right|}{|G_{01}|}\right] \times \left[\max_{A\in\mathcal{I}_R^{K\max}}\left|\sum_{(i,j)\in A}\varepsilon_{i,j}\right|\right] \geq \alpha/2\right)$$

$$+ \mathbb{P}\left(\frac{4}{n(n+1)}\left[\frac{\left|\sum_{(i,j)\in G_{01}}\varepsilon_{i,j}\right|}{|G_{01}|}\right]n_0\overline{\lambda} \geq \alpha/2\right) =: P_1 + P_2.$$

We shall now provide upper bounds for P_1 and P_2.

Let us now provide an upper bound for P_1. Let Z_1 and Z_2 be two non negative random variables, then for all positive α and γ,

$$\mathbb{P}\left(Z_1 Z_2 \geq \alpha\right) \leq \mathbb{P}\left(Z_1 \geq \gamma\right) + \mathbb{P}\left(Z_2 \geq \frac{\alpha}{\gamma}\right).$$

Applying this inequality to P_1 gives

$$P_1 \leq \mathbb{P}\left(\left[\frac{\left|\sum_{(i,j)\in G_{01}} \varepsilon_{i,j}\right|}{|G_{01}|}\right] \geq \gamma\right) + \mathbb{P}\left(\frac{8}{n(n+1)}\left[\max_{A\in\mathcal{I}_R^{K_{\max}}}\left|\sum_{(i,j)\in A}\varepsilon_{i,j}\right|\right] \geq \frac{\alpha}{2\gamma}\right). \tag{51}$$

By Lemma 5, we obtain

$$\mathbb{P}\left(\left[\frac{\left|\sum_{(i,j)\in G_{01}} \varepsilon_{i,j}\right|}{|G_{01}|}\right] \geq \gamma\right) \leq 2e^{-\frac{|G_{01}|\gamma^2}{4\beta}}. \tag{52}$$

Moreover, by Lemma 6, we obtain

$$\mathbb{P}\left(\frac{8}{n(n+1)}\left[\max_{A\in\mathcal{I}_R^{K_{\max}}}\left|\sum_{(i,j)\in A}\varepsilon_{i,j}\right|\right] \geq \frac{\alpha}{2\gamma}\right) \leq 2|\mathcal{I}_R^{K_{\max}}|e^{-\frac{\alpha^2 n(n+1)}{512\gamma^2\beta}} \leq 2C_1 n^{K_{\max}}e^{-\frac{\alpha^2 n(n+1)}{512\gamma^2\beta}}, \tag{53}$$

where we used that $|\mathcal{I}_R^{K_{\max}}| \leq C_1 n^{4K_{\max}}$ (see Equation (50)).

Let us now provide an upper bound for P_2. Using Lemma 5 we obtain the following upper bound:

$$P_2 \leq 2e^{-\frac{|G_{01}|\alpha^2 n^2}{64\bar{\lambda}^2\beta}}, \tag{54}$$

where we used that $n_0 \leq n(n+1)/2$. The proof of Lemma 4 thus follows from (51), (52), (53) and (54).

Technical lemmas.

Lemma 5. *Let A be a subset of $\{1,\ldots n\}^2$. Assume that (A1) holds then for all positive α,*

$$\mathbb{P}\left(\left|\sum_{(i,j)\in A}\varepsilon_{i,j}\right| \geq \alpha\right) \leq 2e^{-\frac{\alpha^2}{4\beta|A|}}.$$

Proof of Lemma 5. By the Markov inequality and (A1), we get for all positive η, that

$$\mathbb{P}\left(\sum_{(i,j)\in A}\varepsilon_{i,j} \geq \alpha\right) \leq \exp(-\eta\alpha + \beta|A|\eta^2).$$

Taking $\eta = \alpha/(2\beta|A|)$ gives

$$\mathbb{P}\left(\sum_{(i,j)\in A}\varepsilon_{i,j} \geq \alpha\right) \leq e^{-\frac{\alpha^2}{4\beta|A|}}.$$

This concludes the proof since the same bound holds when $\varepsilon_{i,j}$ is replaced by $-\varepsilon_{i,j}$. $\qquad\square$

Lemma 6. *Assume that (A1) holds then, for all $\alpha > 0$,*

$$\mathbb{P}\left(\max_{A \in \mathcal{I}} \left| \sum_{(i,j) \in A} \varepsilon_{i,j} \right| \geq \alpha \right) \leq 2|\mathcal{I}|e^{-\frac{\alpha^2}{2n(n+1)\beta}},$$

where A is any subset of $\{(i,j) : 1 \leq i \leq j \leq n\}$ and $\mathcal{I}$ is a collection of any such subsets A.

Proof of Lemma 6. By Lemma 5,

$$\mathbb{P}\left(\max_{A \in \mathcal{I}} \left| \sum_{(i,j) \in A} \varepsilon_{i,j} \right| \geq \alpha \right) \leq \sum_{A \in \mathcal{I}} \mathbb{P}\left(\left| \sum_{(i,j) \in A} \varepsilon_{i,j} \right| \geq \alpha \right) \leq 2 \sum_{A \in \mathcal{I}} e^{-\frac{\alpha^2}{4|A|\beta}} \leq 2|\mathcal{I}|e^{-\frac{\alpha^2}{2n(n+1)\beta}},$$

since $|A| \leq n(n+1)/2$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

AGROPARISTECH/UMR INRA MIA 518

E-mail address: `vincent.brault@agroparistech.fr`